"Models for Planning Capacity Expansion in Local Access Telecommunication Networks"

A. Balakrishnan, T. L. Magnanti
A. Shulman, R. T. Wong

MIT Sloan School Working Paper #3048-89-MS

Revised: August 1989

"Models for Planning Capacity Expansion in Local Access Telecommunication Networks"

A. Balakrishnan, T. L. Magnanti
A. Shulman, R. T. Wong

MIT Sloan School Working Paper #3048-89-MS

Revised: August 1989

Models for Planning Capacity Expansion in Local Access Telecommunication Networks †

A. Balakrishnan *
T. L. Magnanti **
Sloan School of Management
M. I. T.
Cambridge, Massachusetts

A. Shulman
GTE Laboratories, Incorporated
Waltham, Massachusetts

R. T. Wong ***
Krannert Graduate School of Management
Purdue University
West Lafayette, Indiana

May 1989
Revised: August 1989

† This research was initiated through a grant from GTE Laboratories, Incorporated

* Supported in part by an AT&T research award

** Supported in part by Grant # ECS-8316224 from the Systems Theory and Operations Research Program of the National Science Foundation

*** Supported in part by ONR Contract No. N0000-14-86-0689 from the Office of Naval Research

Abstract

The rapid progress of communications technology has created new opportunities for modeling and optimizing the design of local telecommunication systems. The complexity, diversity, and continuous evolution of these networks pose several modeling challenges. In this paper, we present an overview of the local telephone network environment, and discuss possible modeling approaches. In particular, we (i) discuss the engineering characteristics of the network, and introduce terminology that is commonly used in the communications industry and literature; (ii) describe a general local access network planning model and framework, and motivate different possible modeling assumptions; (iii) summarize various existing planning models in the context of this framework; and, (iv) describe some new modeling approaches.

The discussion in this paper is directed both to researchers interested in modeling local telecommunications systems and to planners interested in using such models. Our goal is to present relevant aspects of the engineering environment for the local access telecommunication network, and to discuss the relationship of the engineering issues to the formulation of economic decision models. We indicate how changes in the underlying switching and transmission technology affect the modeling of the local telephone network. We also review various planning issues and discuss possible optimization approaches for treating them.

1. Introduction

Over the last three decades, communication network planning and routing has been a fertile problem domain for developing and applying optimization models. Two main driving forces underlie these modeling efforts: (i) the enormous investments in communication facilities (estimated at around $60 billion in 1980 in Bell System transmission facilities alone (AT&T Bell Laboratories (1986)), and over $100 billion in total for the U. S.) offer significant opportunities for cost savings with even modest improvements in the design and operation of communication networks, and (ii) rapid technological and regulatory changes provide novel design alternatives and operating environments. This paper reviews and develops alternative modeling approaches for addressing contemporary design problems that arise in one major component of a telecommunication system: the local access network. As a starting point, the paper first sets a backdrop for our discussion by reviewing relevant technological developments as well as the evolution of the local access network.

In the next few years, the nature of services and the volume of demand in the telecommunications industry should change radically. Several developments mark the emergence of a new era in communications: replacement of analog transmission by digital technology, decreasing cost and increasing bandwidth of fiber optic transmission equipment relative to conventional copper cables, increasing competition among providers of telecommunication services, and adoption of international Integrated Services Digital Network (ISDN) standards. As ISDN becomes fully operational, and telephone companies complete the transition to digital switching and fiber optic transmission, users will have access to a broad range of new services combining voice, data, graphics, and video. New applications include telemetry, database access, videophone facilities, improved networking services, access to packet networks, and customer controlled network management. Telephone companies are already planning for an even more ambitious expansion of services and capabilities (the so-called broadband ISDN network) when fiber optics will permeate the entire communication system, all the way to the individual customers' homes (Kostas (1984), Dettmer (1985), Toth, Colombini, McClaren, and Yates (1985), *The Economist* (1987), *Fortune* (1988)). Thus, ISDN combined with the new switching and transmission technologies is expected to greatly stimulate network usage.

To accomodate the anticipated demand increase, telephone companies have initiated extensive modeling and planning efforts to expand and upgrade their switching and transmission facilities.

Network modernization and expansion is particularly critical in the local access component of the communication system, both for strategic and economic reasons. In the last few years, the long-distance carriers have almost completed the transition to digital switching technologies and fiber optic transmission. In contrast, the technological changes in the local telephone network, which accounts for approximately 60% of the total investment in communication facilities, have been much more modest. For instance, in 1987, only 20% of all the local access networks in the U. S. employed digital switching (*The Economist*, 1987). Thus, the ability to offer the proposed advanced ISDN telecommunication services is limited by the current capabilities of local networks, and local telephone companies face competitive pressures to upgrade their networks rapidly.

Because modernizing and expanding switching and transmission facilities requires enormous investments, telephone companies typically priortize expansion projects based on demand growth potential and emphasize cost effectiveness in implementing the selected projects. For each project, network planners face complex choices concerning where and when to expand capacity or replace current technology in order to meet the increasing demand for different types of services. The emergence of new communication technologies has created additional decision alternatives and tradeoffs and, hence, new modeling challenges that did not arise in the traditional analog and copper environment. For instance, deploying concentrators and multiplexers in the local access network now provides an alternative method (instead of cable expansion) for increasing network capacity. As a consequence, network planners require new decision support models to identify cost effective expansion and modernization strategies.

This paper focuses on contemporary expansion planning models for the local access component (from the customer premises to the serving switching center) of public telephone networks. We do not address design issues, such as the blocking of potential transmissions or network vulnerability, that are more relevant for long-distance networks. Similarly, our models might not apply directly to data networks or rural networks since these latter network types employ different technologies (for example, rural networks use radio transmission, and data networks employ packet switching) and different design criteria (e.g., reducing packet delay in data networks).

Our purpose in this paper is to discuss alternative modeling approaches rather than a specific methodology for local access network planning. The various models that we consider differ in their

underlying assumptions, complexity and computational tractability. We focus on economic models for aggregate planning (also called *fundamental* planning in the industry) rather than on detailed engineering models of different technologies. Thus, we are concerned with identifying the broad pattern of network evolution, specified by the capacity, location, and timing of investment in different switching and transmission resources. We review some of the underlying telecommunications technology, and contrast the traditional network planning methods developed for the copper and analog environment with the requirements imposed by the newer technologies. We also briefly describe solution methods for the different models.

The rest of this paper is organized as follows. Section 2 describes the evolution and engineering characteristics of local telecommunication networks, and introduces some terminology commonly used in the communications industry and literature. This discussion has two purposes: (i) to highlight technological issues that are important in formulating appropriate optimization models, and (ii) to introduce analysts who might not be familiar with the telecommunication industry to some of the prevailing and expected technology. Section 3 develops a general framework based on a layered network representation that encompasses a wide range of single-period local access network planning models, and motivates different possible modeling assumptions. Section 4 discusses several planning models in the context of our modeling framework. We first review some models proposed in the literature, and then describe two new models - one using a fixed-charge network design formulation, and another based on tree covering concepts. Section 5 offers concluding remarks.

2. The Local Telecommunication Network

This section describes the local access network, traces its evolution over the last few decades, and introduces some communications terminology. Our intent is to describe some important technological features so that we can represent them adequately in economic planning models.

2.1 The Communication Network Hierarchy

Most national telecommunications networks can be broadly divided into the three main levels shown in Figure 1, namely,

(i) the *long-distance, toll* or *inter-city* network that typically connects city pairs through *gateway nodes* (also called *point-of-presence* nodes, Lavin (1987));

(ii) the *inter-office* or *switching center* network within each city that interconnects *switching centers* (also called *local exchanges*, or *central offices*) in different subdivisions (clusters of customers), and provides access to the gateway node(s); and,

(iii) the *local access* network that connects individual subscribers belonging to a cluster to the corresponding switching center.

These three levels of the communication system hierarchy differ in several respects: the processing capabilities and amount of intelligence they contain, the technologies they employ, the services they perform, and their design criteria. For instance, the *local access network* typically has a tree configuration and contains a dedicated communication channel connecting each customer to the switching center. Currently, most (approximately 80% in the U.S.) local access networks use analog transmission on copper cables, and do not contain electronic devices. In contrast, the *long-distance network* has a relatively dense topology providing multiple communication paths between each origin-destination pair. The gateway nodes on this network contain intelligent hardware to perform switching, traffic compression (concentration), and some service functions (such as directory assistance). The long-distance networks in the U.S. are almost completely digitized, and employ high frequency transmission using fiber optics, microwave (radio), and satellite communications. The *inter-office*

network links all the switching centers within a restricted geographical region (for example, within each city) via high speed transmission lines and possibly through tandem switches; it also provides access to the nearest gateway node of the long-distance network. The inter-office network contains limited intelligence for routing incoming messages to the appropriate downstream switching centers or gateway nodes.

Ideally, the design of a telecommunication network should simultaneously account for all three levels of the network hierarchy since the capacity requirements at the different levels are interdependent. For instance, the number of customers assigned to each switching center, and their respective communication requirements (i.e., to whom they communicate and how often) would determine the desired switching capacity at the switching center as well as the transmission capacity on the inter-office network. However, because of differences in ownership and to simplify the planning task, analysts decompose the overall planning problem by considering each level separately (see, for example, Dawson, Murphy, and Wolman (1984)). In this paper we focus on decision models for designing and expanding the local access component.

2.2 Evolution of the Local Access Network

The local access network (also called the *outside plant, local loop,* or *local exchange network*) links individual customers to a switching center that interconnects them for local communications, and also serves as an interface to the higher levels in the network hierarchy. Like the overall communication system, this network is also hierarchical; it consists of three levels referred to as *routes, feeder networks,* and *distribution networks.*

A *route* is a portion of the local access network; it contains all customer nodes that communicate with the switching center via the same link incident to the center. Figure 2 shows a typical route of a local access network. Each switching office might serve as the termination point for 3 to 5 routes (Koontz (1980)); the Bell System contains around 40,000 such routes (Ciesielka and Douglas (1980)). For purposes of capacity planning in the local network, each route is independent of the others.

Each route is in turn divided into two segments: the *feeder network* connecting the switching center to intermediate nodes called *distribution points* (or *control points*), and *distribution networks* connecting each distribution point to the customer premises. The *feeder network* consists of cable groups of varying gauges that are either buried, installed in ducts, or mounted on poles, and are accessible at intermediate points. The segment of cables between two adjacent distribution points along the route is often called a *feeder section*. The feeder network has a tapered structure, i.e., the number of cables in each feeder section decreases as we move away from the switching center. The *distribution network* taps into the feeder network via lateral cables at the distribution points. The area served by a distribution network, sometimes called an *allocation area* (Gibson and Luber (1980)), typically has a diameter of a few thousand feet, and may include as many as 500 customers. The number of distribution points assigned to a switching center varies from 20 to 200. Most feeder and distribution networks have a tree structure and thus provide a unique transmission path from each customer to the switching center. (See Griffiths (1986) for a more comprehensive description of local telecommunication networks.)

Traditionally, the distribution network is designed for ultimate demand (which is relatively small) in order to exploit economies of scale, and to avoid subsequent disruption of service for laying new cables. On the other hand, feeder networks are designed to meet only the medium-term demand; telephone companies periodically review and increase feeder capacity to accomodate demand growth and customer movement (Ciesielka and Long (1980), Elken (1980), Friedenfelds and McLauglin (1979)). In this paper, we focus on the medium-term <u>feeder</u> capacity planning problem. We next trace the evolution of technologies and planning practices in the feeder network. From a modeling perspective, we might classify the technological developments into three stages (see also Dawson et al. (1984)).

<u>Stage 1</u>: *The basic feeder network*

The basic feeder network employs analog transmission at the voice frequency of 4 Khz over copper cables (twisted wire pairs). It uses a dedicated line to connect each customer to the switching center. Physically, the line for a customer might consist of wire segments (possibly with different gauges) belonging to each downstream feeder section, with sections joined at the intermediate distribution points.

In this setting, a principal design concern is to provide acceptable transmission quality by ensuring that the circuit connecting each customer to the switching center satisfies the maximum permissible wire resistance (around 1300 ohms, increasing to 2500 ohms with *range extenders*). Thus, the network engineering task, sometimes called Resistance Design (Ciesielka and Douglas (1980)), is to determine a cost-effective combination of wire gauges to use in each feeder section that will satisfy all maximum resistance requirements.

Observe that the basic feeder network can respond to increased telecommunication demand only by adding and reassigning cables within each feeder section. Planners sometimes refer to this method as *physical pair facility relief*. Any section where demand exceeds the available cable capacity is said to have *exhaust*. The feeder planning exercise considers two strategies to relieve exhaust when customers move or the number of customers increases: (i) feeder cable reallocation, and (ii) feeder cable expansion.

Given the projected changes in the medium term demand at each distribution point, *feeder reallocation* methods attempt to identify a feasible reassignment of currently allocated and spare feeder cables within each section to various upstream distribution points in order to delay cable expansion. Gibson and Luber (1980) describe a heuristic feeder allocation method; Elken (1980) formulates the reallocation task as a separable convex programming problem, and proposes an iterative procedure that solves a sequence of linear programs.

Feeder expansion models (e.g., Freidenfields and McLaughlin (1979) and Koontz (1980)) determine the number of additional cables to install in every time period (typically, every year) of the planning horizon in order to relieve the projected exhaust at minimum total discounted cost. When cable expansion is the only available method for relieving exhaust, and if the local access network has a tree structure, each feeder section can be analyzed independently for capacity planning purposes (since the number of customers served by each distribution point uniquely determines the cable requirements in each section). Thus, physical pair facility relief models used in this context do not incorporate any spatial coupling between sections. Friedenfelds and McLaughlin (1979) formulate a multiperiod capacity expansion model to find the optimal mix of cable gauges in each section; they describe a heuristic solution method for this model.

Stage 2: *Feeder networks with remote electronics*

From a modeling point of view, the next major stage in local network evolution occurred when the communication industry developed *pair gain* or *remote electronic* devices, i.e., *multiplexers, concentrators* and *remote switches,* for use in the local network. A multiplexer is an electronic device that compresses or interleaves signals from several incoming lines into a composite outgoing signal that has a higher frequency but requires only a single line (or a pair of lines). The system assigns each incoming signal to a separate 'channel' in the combined outgoing transmission. (Channels correspond to preassigned non-overlapping frequency bands in frequency division multiplexing, and to time slots in time division multiplexing.) We refer to the ratio of input to output signal *frequencies* as the **traffic compression ratio** (also called the **multiplexing** ratio). Like multiplexers, concentrators also perform traffic compression, transforming multiple incoming signals into a single outgoing high frequency signal. However, the output signal from a concentrator does not have a dedicated channel for each input line (and so signals might be blocked); rather, the output channels are dynamically assigned to the input lines as the need arises. The ratio of incoming to outgoing *channels* is called the **concentration ratio.** Remote switches perform local switching functions to interconnect all customers who communicate with the switching center through them; thus, each remote switch is a decentralized and smaller version of the main switching center. Typically, remote switches also perform multiplexing and/or concentration functions for traffic destined to the main switching center.

In local access network applications, remote electronic devices enable multiple users to share the same physical line on the feeder network, thus providing an alternative method to relieve exhaust as demand increases. They also eliminate circuit resistance restrictions, and permit the use of fewer and less expensive wire gauges. Multiplexers, concentrators, and remote switches are available in several configurations, with varying input capacities (i.e., number of input lines) and different traffic compression ratios ranging from 2:1 to as high as 96:1.

While multiplexing, concentration and remote switching reduce the number of cables required in downstream feeder sections, these cables must now handle higher frequencies. Conventional copper cables (twisted pairs) have a limited effective bandwidth (around 150 Khz). Higher frequency signals (150 Khz to 2 Mhz) require either coaxial cables or *conditioned* (or *groomed)* copper cables, i.e., twisted wire pairs with intermediate repeaters (which are electronic devices to eliminate distortion in the

signals); very high frequency signals (over 2 Mhz) require fiber optic cables. Often, existing ducts (built for copper cables) can accomodate these enhanced transmission media as well.

For the second generation local access network with electronic devices, the planner must consider various choices for locating, sizing, and timing the installation of remote electronics (multiplexers, concentrators, switches) as well as the conventional option of physical pair facility relief (i.e., increasing cable capacities in different feeder sections). Furthermore, unlike the older technologies, these new devices introduce spatial couplings, i.e., we can no longer consider each feeder section in isolation since higher demand at upstream distribution points does not necessarily translate into increased cable capacity requirements on every intermediate feeder section.

Stage 3: *Fiber in the local access network*

Many telephone companies are currently planning to introduce fiber optic technology in local access networks because of its extremely high bandwidth. Fiber optic or lightwave transmission facilities consist of a pair of *fiber optic terminals* (or *fiber terminating equipment*) connected by a fiber cable. Fiber optic terminals convert electrical (analog or digital) signals into very high frequency optical signals, and might perform optical coupling and multiplexing functions as well. The bandwidth of fiber cables (some with a transmission capability of over 1 Tera bits per second) is effectively unlimited for local network applications; therefore, they can accomodate a large number of multiplexed high frequency channels. Indeed, the electronics in the fiber terminating equipment is currently the main factor limiting the number of channels that can be multiplexed on fiber. For local access networks, the cost of fiber terminating equipment is expected to dominate the fiber cable costs (especially, if fiber cables are installed in existing underground ducts) due to the relatively short distances between the distribution points and the switching center.

Except for a few experimental networks, telephone companies have not deployed fiber optic transmission extensively in the local access network. The characteristics, capabilities and deployment of the technology, and even the network configuration plans are constantly changing (see, for instance, Anderson (1988), Carse (1986), Ensdorf, Keller, and Kowal (1988), Toth et al. (1985), Snelling and Kaplan (1984)). Researchers have proposed several competing topologies for fiber-based local access networks (e.g., Campbell (1988), Garbanati and Palladino (1988), Sirbu and Reed (1988), White (1988)).

Some of the topologies, such as the ring, loop, and mesh architectures, seek to reduce network vulnerability, an issue that is becoming increasingly important for fiber networks since damage to even a single cable can affect service to a large number of customers.

In this paper, we assume that, for topological design purposes, fiber optic terminals essentially act like concentrators with very high traffic compression ratios. Thus, we do not represent the unique characteristics of fiber optic transmission in great detail, particularly since this technology is still evolving. When the technology develops further and telephone companies gain experience with deploying it, network planners might require more sophisticated models that distinguish fiber optic transmission from conventional electrical transmission.

Table I summarizes the technologies and expansion options we have discussed in this section. The next section formalizes the feeder network planning problem, and motivates several possible modeling assumptions; we will subsequently use these assumptions to differentiate various modeling approaches.

Table I

Local Access Network Expansion Options

Technology Stages	Technology	How Technology Provides Additional Capacity
1. Basic feeder network	Copper cables	Physical Pair Facility Relief: • Reallocate cables (Feeder Reallocation) • Add cables (Feeder expansion)
2. Feeder network with remote electronics	• Multiplexers • Concentrators • Remote switches	• Compresses or interleaves signals • Compresses signals with dynamic channel assignment • Provides local switching
3. Fiber in local access network	Fiber cables Fiber optic terminals	Converts electrical signals to very high frequency optical signals (with multiplexing)

3. Local Access Network Planning: Problem Definition, Modeling Framework and Assumptions

This section presents a general conceptual and modeling framework for studying the local access (feeder) network planning problem, defines the scope of models that we will subsequently consider, and identifies some key assumptions that distinguish different modeling approaches for this problem. We present an example to clarify the decisions and tradeoffs in the planning task, and to illustrate the modeling implications of traffic processing and cable expansion options. We then develop a general layered network representation that captures the important issues and tradeoffs of the local access network planning problem.

The feeder capacity planning exercise begins with a forecast (based on new construction and customer movements) of telecommunication demand at each distribution point for the duration of the planning horizon. The basic unit of demand for voice transmission is a *circuit*. For analog transmission, each circuit represents a bandwidth requirement of 4 Khz, and requires one twisted pair of copper wires. The corresponding digital equivalent in the U.S. is the DS0 signal which has a transmission rate of 64 Kbps (Kilobits per second). The demand for data, video, and other wideband services is usually expressed as a multiple (or fraction, for some types of data transmission) of the basic DS0 rate; for example, the DS1, DS2, and DS3 rates are, respectively, 24, 96, and 672 times the DS0 transmission rate.

The objective of the planning exercise might be to satisfy all the projected demand at minimum total (investment plus operating) discounted cost, or to selectively satisfy demand to maximize total profit. In this paper we focus on the cost minimization perspective partly because telephone companies currently use this form most often. Furthermore, we restrict our discussions to *single-period* or *static* models rather than *multi-period* models. Multi-period models can account for the temporal couplings caused by economies of scale, i.e., the optimal investment strategy might install excess capacity during one year in anticipation of higher demand in subsequent years. However, multi-period models are much harder to solve than a static model that seeks only to satisfy demand in the terminal year of the planning horizon. Studying static models might possibly give us insights about the more general problem. For instance, the single-period solution algorithms could serve as building blocks for multi-period versions (see, for example, Shulman and Vachani (1988)). Or, as Minoux (1987) has proposed,

the static model might be used to first identify the final target network; a subsequent multi-period model would then determine the evolution of the existing network toward the target.

To meet the demand for different services, the network design must provide adequate and appropriate traffic processing (multiplexing, concentration and switching) and transmission facilities from each distribution point to the switching center. In general, different services and processing devices require different transmission rates and, hence, different transmission media (twisted wire pairs, groomed copper cables, coaxial cables, and fiber optic cables). For example, video signals cannot be transmitted over twisted pair copper cables; similarly, remote electronic devices with high multiplexing ratios require enhanced media that can carry high frequency signals. The local access network design must account for these bandwidth specifications and provide compatible communication resources.

In addition to providing adequate capacity to meet the projected demand, the local access network configuration must also satisfy various technological and policy restrictions. For example, the telephone company might wish to provide multiple paths to some preferred large-volume business customers. Similarly, to ensure adequate transmission quality (particularly for data transmission applications) and to facilitate future expansion and modernization, designers might specify a maximum permissible distance (some companies use a limit of 12 kilofeet) between each customer and the nearest electronic device; we refer to this type of constraint as a *proximity restriction*.

Thus, the single-period local access network planning problem has the following ingredients:

Given (i) the projected (terminal year) demand for different services at each distribution point,
(ii) the current switching and transmission capacities at each location, and
(iii) the costs of installing, expanding and operating new switching and transmission facilities,

find the cost minimizing expansion plan that meets the demand and satisfies all technological and policy restrictions.

The optimal expansion plan should specify (a) the location and size of various network enhancements (i.e., addition or expansion of transmission media and remote electronic devices), and (b) the routing of traffic from each distribution point to the switching center.

This definition of the local access network planning problem makes some implicit assumptions. For instance, we have ignored special *information* processing steps (such as database queries) required by some services. Currently, all information processing occurs at higher levels of the communication hierarchy (i.e., in the inter-office or backbone network); the local access networks perform only *traffic* compression. Future designs might decentralize certain information processing functions as well to the local access network. Also, our single-period model does not consider tradeoffs between overlay and replacement strategies for new technologies. When we consider a multi-period framework for network modernization, the planner must also decide whether to introduce new digital technology by initially overlaying existing analog technology, or by immediately replacing the analog components. Combot and Epstein (1979), Combot and Mason (1979), Combot, Tsui, and Weihmayer (1981), and Hoang and Lau (1984) consider these issues for inter-office network planning.

Before presenting a general modeling framework for the local access network planning problem we discuss a simple example that illustrates the basic decisions and tradeoffs that the network planner must address.

3.1 Example

Figure 3a illustrates a local access network problem with a single service type. The given network has a tree structure, consisting of a single medium (say, copper cables) with existing capacities as shown in the figure. However, this capacity is inadequate for the projected demand level. The heavy shaded lines represent feeder sections with projected exhaust (i.e., capacity shortfall). The projected exhaust represents the amount by which telephone companies would need to expand cable capacities in a conventional physical pair relief strategy (that does not employ traffic processing options). Figure 3b shows a representative cost function for expanding cable capacity along the feeder section between distribution points i and j, consisting of a fixed charge plus a (constant) per unit cost.

In this example, the given network does not have any existing processors. To relieve exhaust, the planner can use a processor with a 10:1 traffic conversion ratio: i.e., the processor compresses the traffic entering on every set of 10 incoming lines onto 1 outgoing line. We assume for simplicity that the

copper cables can transmit both the base rate signal at which the service originates and the processor's compressed output signal. Figure 3c shows an illustrative processor cost function. This cost function might arise if the processor is available in three models with differing costs (G_1, G_2, and G_3) and capacities (Y_1, Y_2, and ∞).

Figure 4 shows one possible expansion plan for the network example of Figure 3. This plan entails installing a processor with a capacity of 400 units at node 5, and expanding the cable segment along sections (3,1) and (4,2) by 100 and 10 units, respectively. The processor at node 5 compresses all the traffic from nodes 2, 4, 5, 8, and 9. Its output signal, shown in dotted lines, travels from node 5 to node 0 (the switching center) via the intermediate nodes 2 and 1; this signal requires only 40 lines since the processor performs a tenfold compression of its 400 incoming circuits. All the other nodes transmit signals at the base (unconcentrated) rate to the switching center. By installing the processor at node 5, we have relieved the projected exhaust on edges (2,5), (2,1), and (1,0) (the physical pair relief strategy would have expanded cable capacities in these feeder sections). Observe that we permit traffic flow in either direction on each edge of the network. For instance, edge (2,5) carries 150 units of unconcentrated traffic (from nodes 2 and 4) from distribution point 2 to the processor located at node 5, and 40 units of concentrated traffic flow in the opposite direction from 5 to 2 (to the switching center); thus, the 200 available lines in section (2,5) can accomodate both these flows. Also note that the expansion plan shown in Figure 4 involves *backfeed*, i.e., flow that is directed away from the switching center, on section (2,5). Some of the models we discuss in Section 4 do not permit backfeed.

The example of Figures 3 and 4 illustrates the two tradeoffs in local access network planning:

(i) *the tradeoff between processors and cable expansion*: Installing a processor at node 5 and assigning nodes 2, 4, 5, 8, and 9 to this processor relieves the exhaust in the downstream sections (5,2), (2,1), and (1,0). This strategy is more economical if the cost of the traffic processor is lower than the cost of expanding cables along the three sections. We could follow a similar strategy to relieve the exhaust on section (3,1). For instance, locating a processor at node 6 (with a capacity of 120 units) to process node 6's traffic would relieve the 100 units exhaust on edge (3,1). However, we might prefer to expand cable capacities if the cost of a 120-unit processor at node 6 exceeds the cable expansion cost (for 100 additional circuits on section (3,1)).

(ii) *the tradeoff between installing a few large centrally located processors versus many small geographically dispersed processors*: For example, should we locate two processors, one at node 4 (with a capacity of 150 units, serving nodes 2 and 4) and the other at node 5 (with a capacity of 250 units, serving nodes 5, 8, and 9), instead of a single processor at node 5? The two-processor solution avoids cable expansion on section (4,2); however, because of economies of scale, its total processor cost is likely to be higher. In general, the total cable expansion cost might possibly increase as the number of processors decreases. On the other hand, installing fewer, but larger, processors reduces the total processing cost. The planning model must address this tradeoff between exploiting economies of scale in processor costs and avoiding transmission capacity expansion by employing a decentralized processor location strategy.

Our example considered a single service, a single processor type, and a single medium that can transport both concentrated and unconcentrated traffic. Additional complexities arise when we consider multiple processor types, multiple processing steps in series, and multiple transmission media. In the next two sections we develop a layered network representation for this more general problem. This representation serves as a framework for comparing various modeling approaches for local access network expansion planning. It encompasses a wide range of existing and proposed transmission and processing technologies, cost structures, and network topologies.

3.2 Modeling Principles

This section discusses the modeling elements for representing local access network planning problems with multiple transmission rates, service types, processor types, and transmission media. In the next section we develop a conceptual framework based on a multi-layer network representation that captures the interrelationship between the three main modeling elements: (i) *customer demands* for different services that must be satisfied; (ii) *transmission facilities* for carrying signals; and (iii) *traffic processors* (e.g., multiplexers, concentrators, remote switches or fiber optic devices) that can compress signals. As mentioned previously, the network expansion plan must match the bandwidth (i.e., transmission rate or frequency) specifications of these three elements. The layered network representation ensures this compatibility by separately identifying the traffic flows at each

transmission rate. We first discuss how the transmission rate serves as a common link between customer demands, transmission facilities, and traffic processors.

We assume that the local access network employs a discrete set of 'standard' transmission rates. For instance, telephone companies in the U. S. use four standard *digital* transmission rates (expressed as multiples of *bits per second* or *bps)* labeled DS0, DS1, DS2, and DS3; these rates correspond, respectively, to 64 Kbps, 1.536 Mbps, 6.144 Mbps, and 43.008 Mbps.

Customer Demands

Customer demands are forecasted by service type (voice, data, video) at each distribution point. Each service type originates at a *basic rate*. For instance, certain types of video services require a basic rate of 1.5 Mbps (the DS1 rate). This rate represents the minimum frequency at which the service must be transmitted; the signals could be multiplexed to higher frequencies, if necessary. The terminal year demand for each type of service is expressed as the number of required channels at the basic rate. The final design should provide the required number of basic rate channels (or an equivalent number of higher rate channels) for each service type from every node to the switching center. Observe that, since we do not consider any unique information processing requirements for different service types, we can effectively aggregate all services requiring the same basic rate into a single service type.

Transmission Facilities

We differentiate transmission facilities according to the medium or cable type (such as twisted wire pairs, copper cables with repeaters, coaxial cables, and fiber optic cables). Each transmission medium can handle a limited set of transmission rates. For example, coaxial cables can transmit DS0 and DS1 signals, while the DS2 and DS3 rates require fiber optic cables. Some transmission media (e.g., copper cables) are typically installed in sections; to connect two non-adjacent nodes the cables must be joined at intermediate distribution points. We refer to these media as *sectional* media. Other types of media (e.g., fiber cables) provide direct *point-to-point* connections without intermediate connectors; however, they use the same physical infrastructure, namely, the underground ducts or trenches in each intermediate feeder section. Most high frequency traffic requires point-to-point cables.

Transmission capacities are specified by the number of lines for each medium between every pair of nodes. In general, the total (transmission) cost to expand these capacities between any pair of node has two components: a ***separable*** cost component pertaining to each individual medium, and a ***joint*** or shared cost that is common to several (or all) media. Joint costs arise because different media share the same infrastructure. For instance, once we incur the costs for building underground ducts, several different media could share these ducts. Because of economies of scale, both the separable and joint cost functions are likely to be concave; for example, they might consist of a fixed cost and a variable cost that depends on the volume of traffic.

Traffic Processors

As explained previously, traffic processors combine several incoming (lower frequency) signals into a single outgoing (higher rate) signal. The economic model that we consider characterize each processor type by three parameters: its **input rate** (i.e., frequency of input signals), **output rate**, and **conversion ratio**. We do not model other detailed technological differences. By *conversion* ratio we mean the ratio of number of incoming *lines* (e.g., copper wire pairs) to outgoing *lines* (including spare lines for contingencies). (The industry uses a related measure called the *pair gain ratio* defined as (number of input lines - number of output lines) ÷ number of output lines.) This conversion ratio may differ from the traffic ***compression*** ratio, i.e., the ratio of output *rate* to input *rate*, because of provisions for spare outgoing lines and differences between the number of input and output channels (as noted previously, concentrators require a smaller number of output channels since they allocate channels dynamically). Effectively, the input and output rates determine the type of input and output transmission media that the processor requires, while the conversion ratio determines the number of physical lines of the output medium required per incoming line.

Processor capacities are specified by their maximum number of input lines. Installing new processors or expanding existing capacities entails *processor costs* which might vary by location, processor type, and the required additional capacity. For instance, this cost function might consist of fixed costs (for acquiring land, constructing buildings, pedestals, cabinets and other infrastructure), and variable or volume-dependent costs (e.g., for each module) that depend on the desired capacity. We expect these processor cost functions to be concave (as a function of additional capacity) because of economies of scale.

A specific commercially available traffic processing device, namely, the SLC-96 system (Ciesielka and Douglas (1980)) illustrates these concepts. The SLC-96 is a modular digital carrier/concentrator system introduced in the Bell System around 1979. Each module supports 96 voice frequency input lines. The input signals are analog; the SLC-96 system converts them into digital signals before retransmission. The input rate of 4 Khz is equivalent to a DS0 digital rate (64 Kbps). The system performs two-to-one digital concentration, i.e., the number of output channels is half the number of input lines; thus, each module has 48 output channels. These 48 output channels are transmitted over two standard so-called T1 digital lines, each carrying 24 channels. The system also requires a spare T1 line to assure continuity of service when one of the main T1 lines fails. Each T1 line might consist of two pairs of copper wires, with intermediate repeaters; the transmission rate over each line is 1.536 Mbps (= 24 channels x DS0 input rate of 64 Kbps), which corresponds exactly to the DS1 rate. Thus, the SLC-96 has a concentration ratio of (96 input channels + 48 output channels) = 2, a traffic compression ratio of (1.536 Mbps + 64 Kbps) = 24, and a conversion ratio of (96 input pairs + (3T1 lines x 2pairs/line)) = 16. One version of the SLC-96 system permits stacking of up to ten modules (to provide service for up to 960 customers).

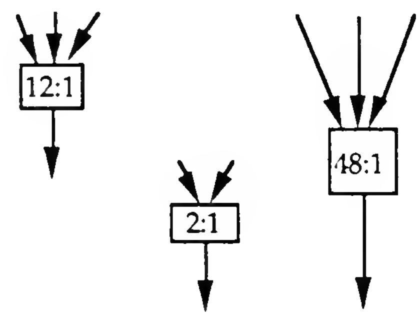
An Illustrative System

Table II summarizes the relationships between transmission rates, customer demands for different service types, transmission facilities, and traffic processors. For ease of illustration, we consider only three transmission rates - DS0, DS1, and DS2 - and assume that each rate has a unique corresponding medium (twisted wire pair, coaxial cable, and fiber cable, respectively). The example considers three service types - voice communication, high speed data, and video service - whose basic rates are, respectively DS0, DS1, and DS2. In this example, the designer can use three types of processors: a type 1 processor to compress DS0 signals to the DS1 rate (i.e., with a DS0 input rate and a DS1 output rate) with a conversion ratio of 12, a type 2 processor to compress DS1 signals to the DS2 rate with a conversion ratio of 2, and a type 3 processor with a conversion ratio of 48 to directly compress DS0 traffic to the DS2 rate. Observe that the conversion ratios differ from the traffic compression ratio; for instance, DS1 signals represent a 24-fold *compression* of the DS0 rate, while the type 1 processor has a *conversion* ratio of only 12 (i.e., it requires 1 outgoing line for every 12 incoming lines). Also, compressing DS0 signals directly using a type 3 processor is more effective (in terms of the numer of

DS2 lines required for the same number of incoming DS0 lines) than using a type 1 and a type 2 processor in tandem; for, say, 480 incoming DS0 lines, the type 3 processor requires only 480/48 = 10 outgoing DS2 lines, whereas the type 1 - type 2 combination requires 480/(12*2) = 20 DS2 lines. The next section presents a layered network representation that models the interrelationships shown in Table II.

Table II
Illustrative System

Layer	Transmission Rate	Service Type	Transmission Medium	Traffic Processors		
				Type 1	*Type 2*	*Type 3*
1	*DS0*	Voice	Twisted pair			
2	*DS1*	Data	Coaxial			
3	*DS2*	Video	Fiber			



3.3 Layered Network Representation

Let G:(N,E) represent the physical network whose nodes N = {0,1,2,...,n} correspond to the switching center (node 0) and the distribution points (nodes 1 to n) asssigned to that center. The set of edges E contains edge (i,j) if the local access network currently contains or can potentially contain a feeder section between nodes i and j. This physical network does not provide a complete representation of the model since it does not differentiate the various transmission rates, transmission media, and processor types. To incorporate these modeling features, we use a multi-layer network that associates a different layer with each transmission rate. Each layer contains a copy of the original network, with

additional edges that represent direct point-to-point cables. The arc flows within a layer correspond to transmission at a specific rate, and flows across layers model traffic compression.

To describe the layered network representation in greater detail, we first consider a representation for the illustrative system summarized in Table II. This example has one service type associated with each transmission rate (DS0, DS1, and DS2), and processor types corresponding to every pair of rates. The main simplifying assumption in this example concerns the available transmission media for each transmission rate. In particular, we assume a one-to-one correspondence between transmission rates and media, i.e., each transmission medium can accomodate only one transmission rate, and every rate requires a unique medium. After developing the layered network for this simplified problem setting, we will indicate network enhancements that will permit us to model transmission media with overlapping frequency ranges.

The layered network, denoted as G_L, contains one layer for each transmission rate. Let us index the transmission rates, and hence the network layers and service types, from $l = 1$ to L in increasing order of frequency. Because of our simplifying assumption concerning the correspondence between transmission rates and media, we can conveniently use the same index l to refer to the transmission medium that carries rate l signals. In the example of Table II, the index $l = 2$ corresponds to DS1 signals (rate 2) as well as high speed data (service type 2) and coaxial cables (medium 2).

Each layer of G_L contains a replica of the nodes of the original (physical) network G:(N,E). We denote the copy of the original node i in layer l as (i,l). The layered network contains two types of edges: *transmission edges* contained within each layer, and *processor edges* connecting the different layers.

The transmission edge connecting node i to node j in layer l, denoted as edge (i,j,l), represents medium l lines connecting distribution points i and j. The flow on this edge corresponds to rate l transmission between i and j. If medium l is sectional, then layer l contains an edge (i,j,l) corresponding to each original feeder section (i,j) in the physical network G. On the other hand, if medium l is point-to-point, then the network in layer l contains edges (i,j,l) for every pair of (original) nodes i and j that medium l can connect. Observe that the edges of this point-to-point network represent the logical, rather than physical, layout of medium l lines; physically, the medium l lines connecting distribution

points i and j would be routed through the ducts of the intermediate feeder sections. Transmission edges are either directed or undirected depending on whether medium l is unidirectional or bidirectional.

The edges connecting two different layers of G_L represent traffic processors. The processor edge connecting node (i,l') in layer l' to node (i,l'') in layer l'' (with $l' < l''$) represents a traffic processor located at distribution point i that compresses rate l' traffic to rate l'' traffic. Thus, for the example of Table II, a processor edge from (i,1) to (i,3) will represent a type 3 traffic processor installed at distribution point i. Observe that processor edges are always directed from lower indexed layers to higher indexed layers since we only permit traffic compression (i.e., transformations from lower to higher transmission rates).

Figure 5 shows the equivalent layered network representation for a local access planning problem that is defined over a (physical) network with a tree topology, and with transmission rates, transmission media, and processor types as shown in Table II. This figure assumes that twisted wire pairs and coaxial cables are sectional, while fiber cables are point-to-point from each potential fiber optic terminal to the switching center. (Thus, layers 1 and 2 have the same topology as the given physical network, while layer 3 has a star topology.)

Our objective is to represent the local access planning problem as a cost minimizing network flow formulation defined over the layered network. To complete the layered network flow formulation, we must specify: (i) the units of demand and flow measurement, (ii) the demands and supplies at different nodes, (iii) the cost functions and capacities of different arcs, and (iv) the laws governing flow conservation at each node. We consider each of these issues in turn.

Units of measurement

When different processor types have arbitrary traffic conversion ratios, we measure the demands, supplies, and flows and capacities of transmission edges in each layer l in terms of the number of medium l lines required. Thus, for the example of Table II, the demand for high speed data is expressed in terms of the number of coaxial lines required for this service at each distribution point. Equivalently, we can use the number of rate l <u>channels</u> as the unit of measurement in layer l since each coaxial cable carries a fixed number of rate l channels (without loss of generality, we can assume that

each rate l line carries one rate l channel). For processor edges, we measure flows and capacities in terms of number of input lines; thus, the capacity of a type 2 processor (in Table II) is specified by the number of incoming coaxial lines. Observe that this measurement scheme implies that the units differ from layer to layer; consequently, as we elaborate later, the problem becomes a network-flow-with-gains formulation rather than a standard minimum cost network flow formulation that conserves flow at each node. In Section 4.3, we consider a special situation in which we can use a common unit of flow measurement for all layers, and transform the general network-flow-with-gains problem to a flow conserving network flow formulation.

Demands and supplies

Each node (i,l) in layer l corresponding to a distribution point i has 'supply' d_{il} equal to the number of medium l lines required to satisfy the requirement for service type l at location i. (Note: For expositional convenience, we treat the line requirements at each node as a supply at that node. Equivalently, we can also define these requirements to be node *demands* rather than supplies.) The switching center node (0,L) in the highest rate layer L serves as the sink (or destination) node for all flows; other switching center nodes $(0,l)$, for $l < L$, serve merely as transshipment nodes.

Edge cost functions and capacities

For representing costs and capacities, we distinguish between four types of arcs: (i) those modeling existing facilities, (ii) those modeling additions to existing capacities, (iii) those modeling new facilities, and (iv) those modeling the collection of traffic at the switching center. To model an existing transmission or processing facility, say, B medium l lines between nodes i and j, we introduce an edge from node i to j in layer l with zero cost but a flow capacity of B. We represent additional transmission and processing facilities by uncapacitated expansion edges (parallel to the capacitated edges representing existing facilities, if any). Similarly, we model new facilities with uncapacitated edges. The cost of flow on the expansion and new facility edges corresponds to the <u>separable</u> cost component for transmission or processor expansion and installation. In addition, as we discussed in Section 3.2, flows might incur <u>joint</u> transmission costs that depend on the combination of flows on all transmission edges (in different layers) that use the same physical feeder section(s). Finally, the cost parameter for the processor edges $(0,l',l")$, for $l' < l"$, incident to the switching center nodes depends on

our assumption regarding permissible transmission rates for traffic entering the switching center. In particular, if the switching center has the capability to receive multiple signal rates, each of these edges has zero cost. If, however, all entering traffic must have the same transmission rate, the processor edge $(0,l',l'')$ carries the cost corresponding to a l'-to-l'' processor located at the switching center.

Flow conservation law

As mentioned previously, because the unit of measurement differs from layer to layer, and flows from different layers interact via the processor edges, we require a network-flow-with-gains equation to relate the incoming and outgoing flows at each node and every layer. To illustrate this constraint, consider some node (i,2) in layer 2 for the example of Table II. This node has three types of incident edges:

(i) transmission edges (i, j, 2) representing coaxial cables carrying DS1 traffic to and from distribution point i;

(ii) an incoming processor edge (i, 1, 2) representing a type 1 processor located at distribution point i that compresses DS0 signals to DS1 signals; and

(iii) an outgoing processor edge (i, 2, 3) representing a type 2 (DS1-to-DS2) processor located at distribution point i.

In addition, distribution point i has d_{i2} incoming DS1 signals corresponding to service type 2 (high speed data) at that location. Each node's incident edges (in the layered network) have associated flow variables which we define as:

$X1_i$ = throughput of the type 1 processor (expressed in terms of the number of incoming DS0 lines) located at node i;

$X2_i$ = throughput of the type 2 processor located at node i (in terms of number of DS1 input lines);

Y_{ji2} = incoming traffic (or line requirements) from j to i on coaxial cables; and,

Y_{ij2} = outgoing traffic from i to j on coaxial cables.

For convenience, we define all four variables to include flow on existing as well as expansion/new facility edges. Observe that the line requirement variables Y_{ij2} and Y_{ji2} are 'directed' even though the edge (i,j,2) is undirected (assuming coaxial cables are bidirectional). $X1_i$, $X2_i$, Y_{ij2}, and Y_{ji2} are the decision variables in the model. Their values would determine how much transmission and processor capacity to add at location i.

The flow conservation law equates the total inflow of traffic at each node of the layered network to the total outflow at that node. The throughput $X1_i$ of the type 1 processor at node (i,2) is one component of the node's inflow; since the throughput of type 1 processors is measured in number of input (medium 1) lines, we must translate this throughput into an equivalent number of medium 2 lines to maintain consistent units of measurement at node (i,2). The conversion ratio, call it ρ_1, of the type 1 processor enables us to express $X1_i$ in equivalent units of medium 2 lines; in particular, the number of medium 2 lines of concentrated traffic arriving at node (i,2) via the type 1 processor is $X1_i/\rho_i$ (which, by the definition of conversion ratio ρ_i, is the number of output lines required by a type 1 processor with $X1_i$ input lines). Thus, the input-output flow equation at node (i,2) has the following form:

$$\sum_j Y_{ji2} + X1_i/\rho_1 + d_{il} = \sum_j Y_{ij2} + X2_i . \qquad (1)$$

To simplify our discussion, we have formulated the flow equation (1) for a specific example, namely, node i in layer 2 for the example shown in Table II. However, the equation readily extends to more general settings in which the node in layer l receives the output of several processor types from one or more lower rate layers $l' < l$, and provides inputs to several processor types that compress layer l traffic to higher rates $l'' > l$.

Observe that equation (1) differs from the flow conservation equation in a standard minimum cost network flow formulation because of the 'loss' factors ρ corresponding to traffic processors. Thus, to model second generation local access networks with remote electronic devices, we require a network-flow-with-gains formulation. In Section 4.3 we show that, under certain assumptions regarding the traffic conversion ratios for different processor types, we can define all decision variables in terms of equivalent base rate channels instead of number of lines for different media. This alternative

definition enables us to use a standard network flow problem formulation that conserves flow at each node.

Extensions

The layered network representation extends to more complex problem settings in which each transmission medium can handle a range of transmission rates (rather than the single rate we assumed in Table II), and the ranges for different media overlap. To represent these overlapping ranges, we could either (i) retain the correspondence between layers and transmission rates, but introduce parallel edges within each layer to model alternative transmission media, or (ii) decompose the problem further by associating a different layer for each transmission rate-medium combination. The first representation is appropriate when we can directly interface two different media at intermediate distribution points without any additional, expensive equipment. For situations requiring interface equipment, the second representation enables us to model the cost of this equipment by associating a cost on interlayer edges that connect two layers corresponding to the same transmission rate but different media. In this case, we also assume that each traffic processing device requires a specific rate-medium combination for both its input and output. Depending on the application context, we might use a mix of these two representations. The layered network can also model different processor technologies as parallel interlayer edges.

Within the framework of this layered network representation, the general local access network planning problem becomes:

Minimize total separable + joint transmission and processor expansion/installation cost

subject to

network flow with gains constraints (1), for all nodes (i,l) of the layered network, and

non-negativity constraints, for all transmission and processor edge throughput variables (X and Y).

Observe that this formulation models multiple processing steps in series, i.e., traffic originating at a node might possibly be compressed at two or more downstream nodes before reaching the switching center. Also, the formulation permits *bifurcated* routes, i.e., two customers connected to

the same distribution point might communicate with the switching center via different routes and processing steps. To prevent this bifurcation, we require a different formulation that distinguishes the traffic originating at different nodes. This latter formulation can also incorporate various proximity restrictions.

Depending on the specific application context, the formulation might contain additional variables and/or constraints. For instance, we can model a fixed plus variable processor cost structure similar to Figure 3b by introducing additional binary variables Zm_i denoting whether ($Zm_i = 1$) or not ($Zm_i = 0$) we install a new type m processor at node i. The fixed cost of the processor serves as the objective function coefficient for variable Zm_i, and the formulation contains additional 'forcing' constraints to relate this location variable to the processor throughput variable Xm_i (see, for instance, Nemhauser and Wolsey (1988)). Similarly, the formulation might contain additional constraints to model certain policy restrictions (e.g., proximity rules, contiguity requirements (see Section 4.3)).

The general network flow formulation with gains and with non-separable edge cost functions is difficult to solve, especially since the problem dimensions might be quite large for practical local access networks. Therefore, models proposed in the literature make several additional assumptions, some that are inherently true for specific application contexts and others that facilitate algorithmic development. These assumptions lead to special problem structures that solution algorithms can exploit. Next, we discuss several categories of possible assumptions. In Section 4 we show how to view various existing and new local access planning models as special cases in the general layered network framework.

3.4 Possible Modeling Assumptions

The problem formulation described in Section 3.3 defines a basic framework for all the local access network planning models that we describe in Section 4. Within this framework, different modeling approaches are possible, each characterized by an additional set of assumptions. These assumptions are motivated by three factors:

(i) They make the model computationally tractable. For instance, certain NP-hard location problems defined over general networks become polynomially solvable for tree networks.

(ii) They reflect uncertainties in the technology.

(iii) They arise because of differences in corporate policies and practices. For instance, some local operating companies might emphasize non-bifurcated routing to reduce the burden of managing/rearranging the network.

We have identified the following six areas that cover most of the differences in modeling assumptions. Selecting different combinations of assumptions in these areas gives rise to different models.

(a) *New versus expansion projects*: Models that apply only to designing new networks (with no existing capacities) are typically easier to solve than expansion planning models that account for existing switching and transmission resources. As we mentioned in Section 3.3, expansion planning models with existing capacities require an additional set of (capacitated) parallel edges, and associated decision variables and constraints that make the model more difficult to solve.

(b) *Network topology*: Some models assume that the physical network has a tree structure. For example, we show in Section 4.3 how this assumption makes a version of the planning model solvable in polynomial time (because tree networks have a unique path from each distribution point to the switching center). Many models assume that all compressed traffic requires point-to-point media connecting the traffic processor directly to the switching center. Effectively, this assumption implies that the network layer corresponding to compressed traffic has a star topology, with the switching center as the central node.

(c) *Backfeed vs unidirectional flows*: Some tree network models incorporate *backfeed*, i.e., traffic movement away from the switching center, while others assume that all flows must be directed toward the switching center. Without backfeed, a processor that is located at node i can serve only upstream distribution points; this restriction limits the set of solutions that the algorithm must consider.

(d) *Processor and transmission cost functions*: In Section 3.3, we discussed generic cost functions for expanding processor and transmission capacity. These cost functions can be specialized in various

ways. First, all the models we discuss in Section 4 ignore *joint* costs, i.e., costs that are shared by two or more transmission media. Further, if all *separable* costs are purely variable, and are linear functions of capacity, we can apply a network flow model (possibly with gains) to solve the local network planning problem. When we include fixed charges for cable and processor expansion, the problem becomes much more difficult to solve.

(e) *Routing restrictions*: Some telephone companies may enforce a non-bifurcated routing restriction specifying that all the traffic from each distribution point must follow the same route (i.e., use the same feeder sections, the same transmission medium on each section, and the same nodal processors at intermediate distribution points). This policy facilitates network management and maintenance. Another possible routing restriction has the following form: if a node contains a traffic processor, then all entering traffic must be processed at that node.

(f) *Single versus multiple processing steps*: Our formulation of Section 3.3 permits multiple traffic processing steps in sequence. If we assume that the traffic from each distribution point can be processed at most (or exactly) once, the number of possible *homing patterns* (i.e., assignment of traffic from various nodes to processors) decreases significantly, thus reducing model complexity. Most existing second generation local access networks employ only one level of traffic processing.

Most local access network planning models can be differentiated along these six dimensions. For example, the tree location model that we consider in Section 4.3 (i) assumes that the physical network has a tree structure, (ii) permits backfeed, (iii) assumes that all enhanced (high frequency) media are point-to-point to the switching center, (iv) prohibits bifurcated routing, and (v) considers at most one processing step for traffic originating at each node. By selecting different combinations of assumptions in this manner we can generate a diverse set of models. In the next section, we outline some possible modeling approaches for the single-period local network planning problem, and differentiate them using this framework.

4. Modeling Approaches for Local Access Network Planning

This section describes some modeling approaches for single-period local access network planning. For each approach, we use the framework of Section 3 to identify the assumptions that differentiate it from other models, and briefly indicate the solution strategy. Sections 4.1 reviews existing models, while Sections 4.2 and 4.3 cover two alternative approaches. Section 4.1 considers the general class of models called concentrator location models, and briefly discusses two other local access network planning methods. In Section 4.2, we discuss how to view, with appropriate assumptions, the layered network model introduced in Section 3 as a fixed-charge network design problem and indicate some possible solution approaches. Section 4.3 describes a more restrictive model that applies only to tree networks; this model can be solved efficiently when designing new networks.

All the models that we review, with the exception of the network design model, consider only a single service type and a single level of traffic processing (the dynamic programming model of Helme, Jack, and Shulman (1988) considers multiple processors in series). Thus, we can represent these models using at most two levels in the layered network (in many, cases we can further reduce the representation to a single layer because of the special cost structures). These models also assume that all high frequency media are point-to-point; thus, each concentrator is directly connected to the switching center in layer 2 of network G_L. Also, all models ignore joint costs, and assume that traffic processors and transmission facilities have separable fixed and/or variable (volume dependent) costs (instead of more general non-linear cost functions). Observe that, when we ignore joint costs and assume a star topology for layer 2 (i.e., each processor is directly connected to the switching center), we can simplify the representation by adding the cost of the point-to-point high frequency medium from each site to the processor cost at that site. Consequently, the local access network planning task reduces to the problem of determining where to locate new processors, and how to connect all the distribution points to the selected processor sites. Most models proposed in the literature apply only to the design of new networks, and do not account for existing facilities.

4.1 Concentrator location and other design models

4.1.1 *Centralized Teleprocessing Design*

In the 1960's and 1970's centralized teleprocessing systems were quite common, and configuring networks to connect users of the system to the central computer was an important design issue (see Boorstyn and Frank (1977), Chandy and Lo (1973), Chandy and Russell (1972), Direlten and Donaldson (1976), Kershenbaum and Boorstyn (1975), Kershenbaum and Chou (1974), Mirzaian (1985), and McGregor and Shen (1977)). These networks typically consist of many (usually 100 or more) geographically dispersed terminals that are connected to a central computer via communication lines. The central computer provides computational resources, and acts as a switch to connect the terminal to a wider distributed computation network. When the number of terminals is large, and the terminals are located in clusters, using concentrators to combine the communications traffic from several terminals (to increase the utilization of communication lines) becomes a cost-effective interconnection strategy. The planning models for these problems typically start with a preselected set of nodes of the network as potential concentrator sites, and select concentrator locations and direct connections from the terminals to the concentrators.

The teleprocessing network design problem consists of three main components: (1) specifying the number and location of concentrators (*concentrator location*), (2) assigning terminals to a concentrator (*terminal assignment*), and (3) determining how to connect every concentrator to its assigned terminals (*terminal layout*). The similarity with the local access network planning problem becomes apparent when we treat terminals as distribution points, and the central computer as the switching center. Most teleprocessing network design methods proposed in the literature first determine the concentrator location and terminal assignment decisions using a single model, called the *Capacitated Concentrator Location Problem* (CCLP), that approximates the actual costs of connecting terminals to concentrators by (separable) assignment costs. Subsequently, a terminal layout method configures the terminal-to-concentrator interconnections based on the assignments suggested by the first phase. (See, for example, Rousset and Cameron (1986).)

We first consider the CCLP as it relates to local access network planning.

Capacitated Concentrator Location Models

Given the fixed installation costs and capacities of new concentrators at each potential site, and the terminal-to-concentrator connection (or assignment) cost, the CCLP selects concentrator locations, and assigns each terminal to one of the selected concentrators in order to minimize the total concentrator and terminal assignment costs, subject to concentrator capacity constraints. The standard CCLP formulation assumes a single servcie type, and provides for only one level of concentration. In terms of our local access network planning framework, CCLP models typically make the following additional assumptions:

(1) *Network topology*: CCLP models effectively assume a double star topology for the final network design: each terminal is directly connected to its assigned concentrator, and each concentrator has a direct connection to the central computer. The equivalent layered network representation contains two layers: the first layer is a complete network with one direct connection from every terminal to each potential concentrator, and the second layer is a star network connecting each potential concentrator site with the switching center.

(2) *Cost structure and capacities*: Essentially, all the CCLP methods proposed in the literature apply only to the design of new networks, i.e., they do not account for existing transmission or cable capacities. Most models incorporate only a fixed cost for each concentrator (that possibly includes the cost of the concentrator-to-computer connection) and each terminal-to-concentrator connection, and assume that the capacity of each new concentrator is prespecified rather than expandable. The total demand for all the terminals assigned to a particular concentrator must not exceed this capacity. Some models specialize the capacity constraint further by assuming that all terminals have equal demand; thus, they limit only the number of terminals assigned to each concentrator.

(3) *Routing restriction*: CCLP models do not permit bifurcated routing. They can accomodate proximity restrictions by setting very high assignment costs for prohibited terminal-to-concentrator assignments.

An enhanced version of the CCLP model, called the *multilevel concentrator location problem* designs a hierarchical structure, where concentrators from one level home on concentrators at the next higher level, and so on.

Even though most CCLP methods implicitly asssume a single concetrator type, we can incorporate multiple concentrator types (each with a different capacity, for instance) by replicating the nodes corresponding to potential concentrator locations, and associating a different concentrator type with each copy. Some models (e.g., Pirkul (1986)) treat concentrator capacities as decision variables by incorporating volume-dependent concentrator costs. While the standard CCLP model assumes that the cost of connecting a terminal to a concentrator does not depend on which other terminals that concentrator serves (by assuming direct terminal-to-concentrator connections), some enhanced models (e.g., Woo and Tang (1973)) indirectly account for the cost economies when terminals are connected by a spanning tree.

Algorithms for the CCLP belong to three broad classes: optimization-based algorithms, heuristic local improvement methods, and clustering techniques.

The CCLP is structurally related to the *plant location* problem, which has been studied extensively in the literature (Cornuejols, Fisher, and Nemhauser (1977), Cornuejols, Nemhauser, and Wolsey (1989), Efroymson and Ray (1972), Erlenkotter (1978), Feldman, Lehrer, and Ray (1966), Kuehn and Hamburger (1963), Sa (1969), Spielberg (1969)). Even though this problem is theoretically intractable, researchers have successfully solved some relatively large-scale uncapacitated plant location problems optimally (Erlenkotter (1978), Geoffrion and Graves (1974), Korkel (1987)). For the CCLP, plants correspond to concentrators and customers to terminals, with the additional restriction on plant capacities. Woo and Tang (1973) propose a CCLP algorithm based on plant location solution methods. Pirkul (1986) proposes an *optimization-based* solution method using Lagrangian relaxation for an enhanced CCLP model that incorporates concentrator sizing decisions (by including variable concentrator costs). The method dualizes the terminal-to-concentrator assignment constraints of a mixed integer programming problem formulation, and iteratively solves single constraint 0-1 knapsack subproblems corresponding to each concentrator location to generate lower bounds on the cost of the original problem. To obtain good lower bounds, a subgradient optimization method iteratively changes the Lagrange multipliers; and, at each iteration, a heuristic procedure uses the Lagrangian subproblem

solution to construct a feasible solution, which provides an upper bound. The author reports computational results for problems with up to 100 nodes and 20 concentrator sites; the percentage gaps between the best upper and lower bounds vary from 0% to 7.7%. For a discussion of other optimization-based and heuristic approaches to the capacitated plant location problem, see Cornuejols, Sridharan, and Thizy (1987).

Polyhedral methods offer a potentially promising approach for solving capacitated plant and concentrator location problems. These methods attempt to refine linear programming (i.e., polyhedral) approximations (relaxations) of these problems by adding (strong) valid inequalities in a cutting plane approach (see, Nemhauser and Wolsey (1988)). Surprisingly little is known about the polyhedral structure of plant and concentrator location problems (for some partial results, see Barany, Van Roy, and Wolsey (1984a,1984b), Ward, Wong, Lemke, and Oudjit (1988), Lemke and Wong (1989) and Leung and Magnanti (1989)).

Local improvement procedures for CCLP start with an initial set of concentrator locations and terminal assignments, and attempt to sequentially decrease the total cost by performing myopic changes. The Add heuristic and the Drop heuristic are two common local improvement methods. The Add algorithm (Kuehn and Hamburger (1963)) is a perturbation method that iteratively evaluates the net savings (savings in terminal assignment costs less the cost of an additional concentrator) that accrue by adding each unused site to the current set of concentrator locations. If no site produces net savings, the method terminates. Otherwise, the most cost-effective site is added to the current set, and the method reevaluates net savings for all remaining sites. Conversely, the Drop algorithm (Feldman et al. (1966)) iteratively removes currently selected concentrators until no further cost reduction is possible. Researchers have also proposed combined methods that alternate between Add and Drop phases. The starting solution for the local improvement procedure can be generated in several ways. One possible initial solution consists of installing a concentrater at every potential location. Rousset and Cameron (1986) have proposed a heuristic to determine the terminal assignments; the procedure first calculates, for each terminal, the difference in assignment costs between the nearest and second nearest concentrators. It then sorts the terminals in decreasing order of this difference and assigns each one, if possible (i.e., subject to the concentrator capacity constraints), to its nearest concentrator.

Several researchers have proposed heuristic methods for the CCLP based on *clustering* concepts (see, for example, McGregor and Shen (1977), Schneider and Zastrow (1982), and Konangi, Aidja, and Dhas (1984)). McGregor and Shen study the single level concentrator location problem, while the other papers address the multilevel problem. Konangi et al.'s method assumes that the terminal-to-concentrator assignment costs are proportional to the Euclidean distance between the two locations, and concentrator costs do not vary significantly by location. For each concentrator level, the method first clusters the terminals into two groups based on geographical proximity, and locates a concentrator at the center (or the potential site closest to the center) of each cluster. Clusters are then successively split if savings result from this splitting, and concentrators are relocated at the centers of the new clusters. When cluster splitting does not give any further savings, a cluster merging procedure attempts to further reduce cost by combining previously defined clusters. After completing the clustering and merging steps for one concentrator level, the algorithm considers the next level, treating the concentrators in the current level as the new terminal locations. The authors report computational results for problems with over 200 terminals.

Clustering methods can also exploit the underlying geometry. For example, under certain circumstances, partitioning the plane into hexagonal cells with a service center located at (or near) the center of each cell is approximately (or asymptotically) optimal. Several researchers have established results of this nature (Fejes-Toth (1959), Fisher and Hochbaum (1980), Haimovich and Magnanti (1988), and Papadimitriou (1981)).

Terminal Layout problem

Given the assignment of terminals to concentrators, the *terminal layout problem* seeks the best network topology connecting each concentrator to its assigned terminals. This model makes the following assumptions:

(1) The original network, containing all available lines, has a general structure; the final topology that is selected must have a tree structure. Effectively, since it does not consider traffic compression, the terminal layout problem is defined over a single layered network.

(2) Each edge of the network carries a fixed charge; the model does not account for variable edge costs, and so effectively ignores cable sizing decisions.

(3) The model incorporates only certain special types of capacity constraints that apply mostly to multidrop lines. (A multidrop line is analagous to a route in the feeder network. It consists of a primary cable emanating from the concentrator to which auxiliary lines attach individual terminals at intermediate points.) These constraints include: (a) *degree constraints*, that specify an upper limit on the number of incident links at a branching node, or at the concentrator, (b) *order constraints*, that restrict the number of intermediate branching nodes between any terminal and the concentrator, and (c) *load constraints*, that limit the total number of terminals that are connected via a multidrop line.

Several authors have proposed heuristic methods for the terminal layout problem (e.g., Esau and Williams (1966), and Sharma (1983)). The Esau-Williams procedure begins with a star network connecting each terminal directly to the concentrator. The method performs a series of edge interchanges to monotonically decrease costs while satisfying the multipoint line capacity restrictions, and terminates when no further cost reduction is possible. A special case of the terminal layout problem that has received considerable attention in the optimization literature is the *capacitated minimal spanning tree problem*. This problem seeks the minimal spanning tree connecting the concentrator to the terminals, subject to degree constraints at the concentrator. Several optimal and optimization-based heuristic algorithms have been proposed for this problem (see, for example, Chandy and Lo (1973), Gavish (1983), Gavish and Altinkemer (1986)). Recently, researchers have begun to study polyhedral approaches (Hall (1989), Hall and Magnanti (1988)) for capacitated spanning tree problems.

4.1.2 Other models

We now describe two other local access network design models, both of which apply only to the design of new networks, i.e., they do not account for existing cable or processor capacities. The first method is a heuristic proposed by Luna, Ziviani, and Cabral (1987) to solve a variant of the local access network planning problem which we call the *service section connection problem*. The second method is a

dynamic programming algorithm developed by Helme, Jack, and Shulman (1988) for a tree network model that prohibits backfeed.

Service Section Connection Model

Luna et al. (1987) consider the following variant of the local access network planning problem. We are given a partition of the set of distribution points into S subsets called *service sections*. Each section contains a number of potential concentrator sites. We must select exactly one site from each service section, and this site will serve all distribution points within the section. The given physical network has contains all permissible interconnections between the potential concentrator sites and the switching center. Each arc of the network carries a fixed cost (for using that arc) as well as a variable cost that depends on the volume of traffic that is routed on that arc; concentrators have fixed costs that might vary by location. The planning problem consists of: (i) selecting one concentrator site from each service section, and (ii) designing a subnetwork that connects all the selected concentrator sites to the switching center. The objective of this service section connection model is to minimize the total (fixed + variable) arc costs plus the concentrator costs. Unlike the CCLP, the service section connection problem explicitly considers the topological design decisions for connecting concentrators to the switching center. However, the model ignores the interconnections within each service section, i.e., it does not consider the topological design decisions for the subnetwork connecting the distribution points within a service section to the selected concentrator site in that section. Effectively, it assumes that, for each potential concentrator site, the designer has predetermined the distribution point-to-concentrator connections; the cost of this subnetwork may be incorporated in the concentrator cost for that site. The model does not consider multiple services, multiple concentrator types or different transmission media, and it does not model economies of scale in concentrator and transmission costs.

The layered network representation of the service section connection problem consists of a single layer containing an augmented version of the given physical network. A single layer representation suffices because the model considers only the topological design decisions for compressed traffic (and ignores the flow pattern for the base rate traffic). For the layered network representation, we augment the physical network as follows:

For each service section, we add a *super* node whose demand equals the sum of the demands for all distribution points in that service section. We connect each super node

to every potential concentrator site in the corresponding section using an edge directed away from the super node. The fixed cost of this edge equals the cost of the concentrator. All the edges of the given physical network (interconnecting potential concentrator sites and the switching center) carry the fixed and variable costs specified in the original problem; all original nodes (i.e., potential concentrator sites) serve as transshipment points. The switching center node has demand equal to the total line requirements in all service sections.

Since the model considers only new facilities without any capacity constraints, it is possible to show that some optimal solution to the layered network problem assigns each super node to exactly one concentrator site in the corresponding service section; thus, we do not need explicit constraints to prevent the model from selecting multiple concentrator sites in each service section. Also, the optimal topology of the service section interconnection network will have a tree structure, and so will not bifurcate any traffic flow. Because of the special (fixed plus variable) cost structure, this model is a special case of the fixed-charge network design model that we present in Section 4.2.

Observe that this model can implicitly incorporate proximity restrictions, i.e., we can choose the service sections so that distribution points that do not meet the proximity criterion with respect to a potential concentrator site do not belong to the same service section as this site. Also, if we interpret the nodes of the network as individual customer locations, the service sections as allocation areas (see Section 2.2), the potential sites within each section as potential distribution points, and the switching center as a concentrator, then Luna et al.'s model applies to the design of the distribution network for each concentrator, rather than the feeder network serving the switching center.

Luna et al. first formulate the service section connection problem as a mixed integer program, and propose a heuristic method to solve it. The heuristic starts by constructing the following design: for each service section, select the distribution point that is closest to the switching center; and, find the shortest path tree (using only the variable arc costs) connecting the switching center to each selected concentrator site. A local improvement procedure then attempts to improve this initial feasible solution. The method iteratively evaluates the cost savings as it interchanges concentrator sites or arcs; it performs profitable interchanges sequentially until no further savings result. The authors report computational results for 3 problems ranging in size from 18 nodes, 54 arcs, and 7 service sections, to 263

nodes, 752 arcs, and 117 service sections. However, they do not evaluate the quality of these solutions since the method does not generate any lower bounds or alternative heuristic solutions.

Tree Network Model without backfeed

Helme et al. (1988) propose a dynamic programming method for a special class of local access network planning problems. The method permits multiple processors in series, but assumes a single transmission medium and applies only to the design of new networks. It assumes that the given network has a tree structure, does not permit backfeed, and does not account for economies of scale. Each processsor has a fixed cost (that may vary by location); transmission facilities have only variable costs. The solution method is based on a recursive procedure that exploits the tree structure. For each node of the network, the recursive relationship determines the cost of connecting that node to the switching center, for various possible combinations of downstream processor locations.

Helme et al. also propose a Drop/Add heuristic for a more general local access network planning model that considers general network topologies and multiple processor types in series, permits bidirectional transmission (i.e., backfeed) on links, and incorporates existing capacities as well as fixed and volume-dependent processor costs. The method ignores fixed costs for cable expansion; it assumes that cable costs are directly proportional to the number of (additional) cables required. The Drop heuristic starts with an initial design containing all processor types at each node, and successively eliminates processors to reduce the total cost. Because of the linear cable expansion cost structure, the authors are effectively able to use a shortest path algorithm to compute the total transmission cost for any given processor configuration.

4.2 Network Design Model

The general layered network framework introduced in Section 3 can be specialized so that it takes the form of the fixed-charge network design problem which arises in a variety of distribution planning, manufacturing and telecommunications contexts. Given the demand between various origin-destination pairs, and fixed and variable costs for each arc of a network, the (fixed charge) network design problem involves selecting a subset of arcs, and routing the various commodities (subject to

conservation of flow, without gains or losses, at each node) over the selected arcs in order to minimize the total fixed plus variable arc costs. The capacitated version of this problem accounts for arc capacity constraints as well. The network design problem generalizes several well-known optimization models including the plant location, shortest path, Steiner tree, traveling salesman, and minimal spanning tree problems. Magnanti and Wong (1984) and Minoux (1989) describe various applications and solution methods for the network design model.

In this section, we demonstrate how, in certain circumstances, to cast the general layered network framework for the local access network planning problem as a fixed charge network design model.

The first assumption concerns the transmission and processor cost structures. The network design model ignores joint costs between various transmission media, and assumes, as in the example of Table II, a one-to-one correspondence between transmission rates and media, i.e., the planner preselects a preferred transmission medium for each transmission rate. We also assume that all processor and cable installation/expansion costs are piecewise linear, consisting of fixed and variable components.

The network design model's second assumption concerns the conversion ratios for different processor types. Recall from our discussions in Section 3.1 that, in general, each processor requires a certain input rate (or frequency), transmits output at a higher rate, and has a specified conversion ratio ρ (defined as the ratio of input to output lines). To develop the network design model formulation, we assume that the conversion ratios for different processor types are *compatible* in the following sense. Consider three processor types labeled 1, 2, and 3, and suppose we can compress rate l_a traffic to rate l_b either by employing a type 1 and type 2 processor in series (i.e., the type 2 processor has the same input rate as type 1's output rate), or by using a single type 3 processor (with input rate l_a and output rate l_b). The *conversion ratio compatibility assumption* requires that the three conversion ratios must satisfy the following equation:

$$\rho_3 = \rho_1 * \rho_2,$$

where ρ_m denotes the conversion ratio for a type m processor. In other words, the messages on x lines in the rate l_a medium always require exactly $(x \div \rho_3)$ lines in the rate l_b medium, regardless of whether the compression was achieved using a type 1 and type 2 processor in tandem, or a single type 3 processor. Effectively, this assumption permits us to associate a single *conversion factor*, call it δ_l, with each

layer l of the multi-layer network. We define this factor as the number of channels (or circuits or lines) at the *base rate* (the lowest transmission rate, corresponding to the index $l = 1$) that each type l line can accommodate. In the example presented in Table II, the three processors do not satisfy the compatibility assumption since the product of the conversion ratios for processors 1 and 2 ($\rho_1 * \rho_2 = 12 \times 2$) does not equal the conversion ratio ($\rho_3 = 48$) of the type 3 processor. The single conversion factor δ_l for each transmission rate enables us to measure all the traffic in every layer in terms of the <u>number of equivalent base rate channels</u> (rather than the number of lines of the corresponding medium). This common traffic measurement unit preserves conservation of flow at each node, i.e., we do not require the conversion factor ρ in the left-hand side of constraint (1) in Section 3.3. Thus, with the conversion ratio compatibility assumption, we can transform the network flow with gains flow equation (1) to a standard network flow conservation constraint.

Apart from these two assumptions, the network design model incorporates all other features of the general layered network framework described in Section 3.3. In particular, it can handle general network topologies, multiple service types (as long as these do not impose unique processing requirements), sectional and point-to-point cable types, economies of scale in processor and transmission cost functions, and existing transmission and processor capacities. It also permits backfeed and bifurcated routing. If the cost functions are piecewise linear and concave, and if the network does not contain any existing capacities, the model reduces to an uncapacitated network design problem. Existing resources and non-concave cost functions introduce arc capacities. We first describe the cost parameters for the uncapacitated fixed charge network design model with no existing processor and transmission capacities, and with a fixed plus linear cost structure for each transmission and processing facility. Subsequently, we discuss extensions to model general piecewise linear cost functions and existing capacities.

<u>Uncapacitated Network Design Model</u>

The transmission and processing costs are represented as edge cost functions in the layered network. Let $H_{il'l''}$ be the fixed cost of a processor located at node i that converts layer l' signals to layer l'' signals, and let $v_{il'l''}$ be its variable cost (per unit capacity, expressed in base rate channels). Thus, for a processor with capacity of x (base rate) units, the total cost is $H_{il'l''} + v_{il'l''} * x$. We associate these fixed and variable costs with the processor edge (i,l',l'') connecting layers l' and l''.

(Note that a per unit cost of $v_{il'l''}$ for the processor located at node i implies that the processor cost per input (medium l) line is $\delta_l * v_{il'l''}$.) Similarly, let F_{ijl} and c_{ijl} represent, respectively, the fixed and per unit costs for a medium l (sectional or point-to-point) connection from node i to node j. (Again, a per unit cost of c_{ijl} implies that each additional medium l line from i to j costs $\delta_l * c_{ijl}$.) These two transmission cost parameters define the fixed and variable costs for the transmission edge (i,j,l). As before, the inter-layer edges connecting the switching center nodes (0,l') and (0,l'') carry zero cost, assuming that multiple signal rates can enter the switching center. Otherwise (if all entering traffic must have the same transmission rate), the processor edge (0,l',l'') carries the fixed and variable costs corresponding to a l'-to-l'' processor located at the switching center.

With this set of model parameters, the uncapacitated network design solution that conserves flow at each node, and satisfies all demands at minimum total fixed plus flow costs corresponds to the optimal local access network plan. In the optimal network design solution, a flow of x_{ijl} units along transmission edge (i,j,l) implies that the number of medium l lines to install in feeder section (i,j) is x_{ijl}/δ_l. Similarly, the flow on processor edge (i,l',l'') divided by $\delta_{l'}$ gives the capacity (in terms of number of input lines) of a processor at node i that transforms layer l' input signals to layer l'' output signals. Observe that the transmission edges might be either directed or undirected, and the model permits multiple processors in series.

It is possible to enrich this network design model in various ways. For instance, we can model economies of scale in processor and transmission costs if these economies can be approximated by piecewise linear, concave cost functions as shown in Figure 6. Suppose the function shown in this figure describes the cost of installing a medium l cable on feeder section (i,j). In general, this cost function contains R breakpoints. Breakpoint r occurs at a capacity of B_r, and the slope of the cost function decreases from c_r to c_{r+1} at this point. Let F_r and F_{r+1} denote the y-intercepts of the two line segments that define breakpoint r. We can incorporate this cost function in the network design model by introducing R parallel arcs between nodes i and j in layer l. The r^{th} parallel arc carries a fixed charge of F_r and a variable cost of c_r. Because the overall transmission cost function is concave, the optimal solution will automatically select the appropriate line segment even without any explicit capacity constraints, i.e., if the optimal local access network solution entails installing a capacity of x units on arc (i,j), and x lies between B_r and B_{r+1}, the network design solution will route all x units on the r^{th} edge since this edge minimizes total cost, among all parallel edges, for the x units of flow. We can

similarly model piecewise linear, concave processor cost functions by introducing parallel processor edges. This type of cost function might represent the composition of alternative multiplexing or concentration technologies with progressively increasing fixed costs but decreasing variable costs. As mentioned in Section 3, the layered network formulation does not readily accommodate proximity restrictions.

Certain properties of the optimal uncapacitated network design solution (with concave edge cost functions and no existing capacities) have special significance for the local access network planning problem. For instance, it is easy to show that, in some optimal network design solution, each node (i,l) either processes all incoming traffic or routes all the traffic to another node on the same level, but not both. Thus, we cannot have both type l traffic leaving node i, and a l-to-l' processor located at node i. In particular, consider the route for, say, layer 1 traffic originating at node i. Let node j be the first node on this route containing, say, a 1-to-l processor. Then, the traffic originating at node i and all intermediate nodes from i to j also undergoes 1-to-l processing at node j. As a corollary, the problem has an optimal solution that does not use bifurcated routing, and satisfies the so-called contiguity property that we define in Section 4.3.

Capacitated Network Design Model

The uncapacitated network design model applies to local access design problems with no existing capacities. Modeling networks with existing processor and transmission capacities, or with piecewise linear but non-concave cost functions requires explicit capacities on the edges of the layered network. In particular, suppose the network already contains x type l lines connecting nodes i and j. To represent this capacity, we add a parallel arc connecting nodes (i,l) and (j,l) (in layer l); this arc has zero fixed and variable costs, but has a capacity of $x^*\delta_l$ (basic traffic) units. Similarly, suppose the cost for expanding the type l transmission facilities in section (i,j) is a piecewise linear and convex cost as shown in Figure 7a. Figure 7b shows the equivalent network representation with parallel arcs and appropriate arc fixed costs, variable costs, and capacities. Figure 8a shows a more general cost structure, and Figure 8b gives its representation.

As we have seen, the specialization of the layered network framework as a fixed charge network design formulation for local access network planning is very versatile. Its assumptions are less

restrictive than previous local access network models, and indeed it generalizes many of the previous models. Next, we briefly outline solution methods for the network design problem.

Solution Methods for the Network Design Problem

The network design problem and several of its variants are known to be NP-hard (Johnson, Lenstra and Rinnooy Kan (1978)). Several authors have proposed heuristic and optimization-based methods for solving the problem (Hoang (1973), Boyce, Farhi, and Weischedel (1973), Billheimer and Gray (1973), Boffey and Hinxman (1979), Dionne and Florian (1979)). Balakrishnan, Magnanti, and Wong (1989a) propose a dual ascent method that generates provably near-optimal solutions to the uncapacitated network design problem. The method essentially approximately solves the dual of the linear programming relaxation of the network design integer programming formulation. The dual solution generates a starting design for a local improvement heuristic, and also provides a lower bound that can be used to verify the quality of the heuristic solutions. This method generalizes several previous algorithms for special cases of the network design problem, including the Steiner tree and plant location problems. The approach was successfully tested on several randomly generated problems containing up to 45 nodes and 595 arcs.

Capacitated network design problems are much harder to solve than the uncapacitated version. One possible solution strategy is to dualize the arc capacity constraints (i.e., multiply the capacity constraints by Lagrange multipliers, and add these multiples to the objective function). The resulting subproblem is an uncapacitated network design problem that can be solved at least approximately using, say, the dual ascent algorithm. By iteratively modifying the Lagrange multipliers using a method such as subgradient optimization (see, for example, Fisher (1981)), we can possibly generate good heuristic solutions and lower bounds for the original capacitated problem. However, previous experience with this approach for the capacitated plant location, capacitated minimal spanning tree, and other related models suggests that the addition of arc capacities significantly increases the gaps between the upper and lower bounds. Thus, we expect local access planning problems with existing capacities and non-concave cost functions to be computationally more difficult.

One of the main limitations of the network design model is its size. In particular, the number of nodes and arcs in the layered network grows very rapidly with the number of different transmission

rates, processor types, and distribution points. To model a problem involving 20 distribution points with all possible connections, 5 processor types, and 3 transmission media, requires a network with 63 nodes, and over 600 (undirected) edges. These problem dimensions probably represent the largest size that current optimization-based network design algorithms can solve within a reasonable amount of computational time. In the next section we describe an alternative specialized model that is more tractable since it assumes a tree network, and imposes certain restrictions on the way distribution points are assigned to concentrators.

4.3 Tree Covering Model

In this section we describe a special case of the local access network planning problem, which we call the *tree covering model,* that is solvable in polynomial time when the network does not contain any existing capacities. The model assumes that the network defining the permissible interconnections is a tree. It also makes some additional assumptions regarding the cost structure and routing policy. The model does permit backfeed, and can incorporate economies of scale. We first describe the model as it applies to the design of new networks, and subsequently describe an enhancement to account for existing capacities.

We say that a node i *homes* on another node j if the traffic from distribution point i is processed at node j. Node i homes on the switching center (node 0) if its traffic is not processed at any intermediate node. The tree covering model makes the following assumptions:

(1) The given physical network, containing all permissible cable sections, has a tree structure, rooted at the switching center. Thus, a unique path connects each distribution point to the switching center.

(2) The model permits at most one level of traffic processing and assumes a single service type. For simplicity, we assume that traffic can arrive at different frequencies at the switching center.

(3) *Contiguity assumption*: The model assumes that if a node i homes on node j, then all intermediate nodes lying on the (unique) path between nodes i and j also home on node j. In particular, node j must home on itself if it contains a processor. We refer to this routing restriction as the contiguity

assumption since the set of all nodes homing on a particular processor induces a single contiguous or connected subgraph of the original network.

(4) The model does not permit bifurcated routing, i.e., all the traffic originating at a particular node must follow the same route to the switching center (i.e., must use the same links, and undergo processing at the same node).

(5) The model permits multiple processor types and transmission media. However, it ignores any joint costs between media, and assumes that all high frequency media are point-to-point to the switching center. This assumption essentially permits us to include in the processor cost all transmission costs for traffic emanating from each traffic processor. The base rate medium (say, twisted wire pairs), which we will refer to as *cables*, is assumed to be sectional.

(6) Each processor type is assumed to have a fixed plus variable cost structure (including the transmission cost for the processor's output) that may vary by location. Similarly, cable installation and expansion entails a fixed and variable cost that varies by section. Like the previous network design model, the tree covering model can also accommodate piecewise linear, concave cost functions.

(7) The model can also account for additional *homing costs* when node i homes on a processor located at node j. By selectively setting these homing costs to a high value, we can prohibit homing patterns that violate proximity restrictions.

Without the additional homing costs, the tree covering model has an equivalent single-layer fixed-charge network representation with some additional routing restrictions. The single layer consists of the original physical tree network with the following enhancements: each node of the tree is connected by a directed arc (directed away from the node) to the switching center. This arc models the traffic processor at the node as well as the point-to-point cable connecting this processor to the switching center. Every node of the network (except the switching center node) has supply equal to the projected number of channels required for the corresponding distribution point; the switching center node is the sink for all flows. The contiguity assumptions translates into a routing restriction that specifies

that each node (except the switching center) has outgoing flow (either at the base rate or at the compressed rate) on exactly one arc.

The tree covering model's assumptions permits us to transform the local access network planning task into a problem of covering all the nodes of the original tree by subtrees. Consider a local access network solution in which node j contains a processor. Let N(j) be the set of nodes that home on this processor, and let T(j) be the subgraph induced by this node subset (i.e., T(j) contains edge (p,q) of the original tree network if nodes p and q both belong to the node subset N(j)). Our contiguity assumption implies that T(j) must be a single connected component, i.e., it must be a subtree of the original tree. Thus, the union of the induced subtrees corresponding to each processor must span all the nodes of the network. (By convention, the switching center always contains a processor.) Conversely, suppose we are given a subtree T that must be served by a single processor located at one of the nodes within the subtree. For each potential processor location in the subtree, we can calculate the exact flows, and hence the exact value of the cable expansion costs for all subtree edges. The processor cost is also known since this processor must serve the total line requirements for all nodes in the subtree. Consequently, we can easily calculate the total transmission plus processing cost of serving all the nodes in subtree T from each node $i \in T$. The node i that minimizes total costs is the best processor location for this subtree.

These properties enable us to solve the tree covering model (without existing cable and processor capacities) very efficiently using a dynamic programming algorithm based on a method developed by Barany, Edmonds, and Wolsey (1986) for optimally covering a tree with subtrees. This method is also closely related to the p-median algorithm discussed by Kariv and Hakimi (1979). The algorithm starts from the leaves of the original tree, and recursively builds the covering solution for successively larger subtrees. Balakrishnan, Magnanti, and Wong (1989b) describe this method in greater detail. Next we outline a different method, using a shortest path algorithm, to solve the special case when the given network is a line network.

A *line network* consists of a simple path connecting two end nodes, one of which is the switching center node, say, node 0. Without loss of generality, assume that the nodes are indexed sequentially from 0 to n so that the line network contains only (undirected) edges of the form (i-1,i), for i = 1,2,...,n. Figure 9 shows this structure. Suppose we want to locate p processors on this network, including the processor at the switching center. Then, by our contiguity assumption, each processor induces a line

segment containing all the nodes it serves, and the union of all p line segments cover all the nodes of the network. Thus, the local access network planning problem for this special case reduces to the problem of determining the number of processors (p) to locate, and the optimal partition of the original line network into p segments. We can formulate this problem as a shortest path problem in the following way. Consider a line segment from node i to node j (inclusive), with j > i. As mentioned previously, we can easily determine the optimal total cost of serving all the nodes in this segment by enumerating all potential processor locations between i and j. For instance, consider a potential location k, and suppose k > i+1. Then arc (i,i+1) must carry node i's supply d_i, arc (i+1,i+2) must carry the cumulative supply for nodes i and (i+1), i.e., $(d_i + d_{i+1})$ and so on. Thus, we can determine the flow on each edge of the i-to-j line segment. Also, the total processor throughput is the sum of the supplies for all nodes from i to j. Using this information, we can determine the total processor plus cable cost for serving all the nodes between i and j, assuming that the processor is located at node k. Let k_{ij} be the best processor location (i.e., the node on the line segment that minimizes total cost) for serving line segment i-to-j, and let c_{ij} be the corresponding optimal cost.

To find the best partition of the original network, we construct a shortest path network defined over the (n+1) nodes. For every $i \leq j$, this network contains a directed arc from j to (i-1). The cost of this arc is set equal to c_{ij}, the optimal cost of serving all nodes between i and j (inclusive). Then, every path from node n to node 0 defines a partition of the line network. In particular, including arc (i-1,j) in the n-to-0 path corresponds to selecting the line segment from node i to j as one element of the partition. Consequently, the shortest n-to-0 path defines the optimal partition of the original network, and hence identifies the optimal local access network configuration. Because of the special structure of the line network problem, the costs c_{ij} can be computed efficiently. Also, the algorithm can easily accomodate existing cable and processor capacities. For situations without backfeed this approach can be simplified even further and is analagous to the well-known Wagner-Whitin model of production planning (Kubat (1985)).

For general tree networks, incorporating existing cable and processor capacities considerably complicates the model and its solution. Balakrishnan et al. (1989b) describe a Lagrangian-relaxation approach that formulates the local access network planning problem as a mixed integer program, and dualizes the capacity constraints. The resulting subproblem is an uncapacitated local access network planning problem that can be solved efficiently using the dynamic programming procedure we

mentioned previously. This solution gives a lower bound on the total cost of the original problem. The authors embed the Lagrangian subproblem solution method in an iterative procedure that modifies the Lagrange multipliers in order to improve the lower bound. The subproblem solution at each iteration also identifies a feasible network expansion plan, which can be improved heuristically to generate good upper bounds. The authors describe various formulation and algorithmic enhancements that significantly improve the method's performance, and report computational results based on some actual test networks.

In conclusion, this section has described various models for local access network planning. We have seen several possible combinations of assumptions, each defining a separate model. Table III summarizes the main differences in assumptions, features, and solution methods for the models that we reviewed. The fixed-charge network design model is quite comprehensive, and recent advances in network design algorithms make this modeling approach computationally feasible for medium-sized local access network planning problems, especially for designing new networks. On the other hand, the general network design model does not exploit any special structure that specific application contexts might possess. For instance, in order to simplify the task of managing the network, some local telephone companies might adopt routing policies that validate the contiguity assumption. Similarly, if each concentrator requires an umbilical (direct) connection to the switching center, ignoring shared media costs and assuming a point-to-point medium for compressed traffic might be appropriate, especially if these enhanced cables can be installed in existing ducts. Making these simplifying assumptions enables us to use specialized algorithms, thus increasing the range of problem sizes that can be solved.

Table III
Local Access Network Planning Models

Model Type	Topology of Physical Network	New or Existing Networks?	Cost Structure	Restrictions	Features	Solution Method
Centralized teleprocessing design model				Point-to-point connection from concentrator to switching center; Single service type		
• Capacitated concentrator location	General	New	Fixed only	No bifurcated routing	Selected design has double star topology	Optimization-based and local improvement heuristics
• Terminal layout	General	New	Fixed only	Only a single cable size can be installed	Terminal layout model usually applied with fixed concentrator locations	Optimization-based approaches
Switching center connection model [Luna et al. (1987)]	General	New	Fixed only	Prespecified service sections, potential concentrator sites in each section, and routing from distribution points to concentrator sites; Single service type	Can incorporate proximity restrictions; exactly one concentrator site selected for each service section	Shortest path/local improvement heuristic
Tree network model without backfeed [Helme et al. (1988)]	Tree	Both	Fixed plus variable	No backfeed	Can accommodate multiple processors in series	Dynamic programming
Network design model with layered network	General	Both	Fixed plus variable	Fixed and (linear) variable processor and cable costs; Conversion ratios for different processor types are compatible	Can accommodate multiple customer services and multiple processors in series; Cannot accommodate customer proximity restrictions	Dual ascent
Tree covering model	Tree	Both	Fixed plus variable	Contiguity assumptions; No bifurcated routing	Can incorporate proximity restrictions	Polynomial solvable for new networks; Lagrangian relaxation/ dynamic programming

5. Concluding Remarks

This paper has attempted to (i) set a backdrop for economic modeling of local access telecommunication systems by tracing the evolution of relevant technology in public telecommunication networks, (ii) provide a general framework for viewing planning and design models, (iii) summarize previous contributions from the literature, and (iv) describe some new modeling approaches. Several technological, economic, and social forces are fueling interest in these economic models, particularly, the rising demand for a variety of services due to the introduction of ISDN standards, the installation of digital and fiber optic technology, and mounting competitive pressures. The traditional local network planning tools are inadequate in the current environment because the availability of electronic traffic compression devices for the feeder network has created new ways to respond to increasing demand for telecommunication services.

For our review of planning models, we focused on static (single period) models, emphasized the differences in modeling assumptions, and briefly outlined solution methods. As our discussion of modeling approaches has suggested, the general area of local access network planning continues to provide many challenging opportunities for modeling and algorithmic development, particularly for multiperiod and multiple service versions of the problem. The new developments in telecommunication standards and technologies should further stimulate the development of new modeling approaches.

Some of the static models offer potential for use in multiperiod settings. For instance, we might employ a decomposition method such as Lagrangian relaxation to decompose the multiperiod problem into several single period problems, each that could be solved using one of the single-period methods. In this scheme, the Lagrangian multipliers corresponding to a time period t might represent the 'price' that we are willing to pay to establish excess transmission and switching resources at time t for use in future periods. Thus, the pricing mechanism accounts for the temporal coupling of plans by acting as an incentive to exploit economies of scale. An alternative approach is to use static models to generate a final target network; we might then apply a different model to plan the evolution, over the multiple time periods of the planning horizon, from the current network to the target network. Shulman and Vachani (1988) propose a related approach.

The models that we discussed did not include any special representation for fiber optic facilities partly because their current economic implications are comparable to those of other electronic traffic processing devices and high frequency media. Future developments and implementation of fiber optics in the local loop might necessitate other distinctions in modeling fiber optic technologies.

The continuing evolution of local access network technology and ever increasing efforts to formulate new ways of utilizing this technology create a number of exciting and challenging future research directions. One promising technological development is the possibility of installing remote switches and other 'intelligent' hardware in the feeder and distribution networks. These devices can perform a number of switching center functions, and in many applications customers would need to communicate only with a nearby remote switch instead of connecting all the way to the switching center. This strategy of using remote switches would reduce the overall traffic in the feeder network, and thus reduce the need for additional cables or processors. Strategies for the proper deployment of these remote switches might be a fruitful topic for future studies.

The enormous bandwidth of fiber optic networks creates intriguing opportunities for developing new services for households: for example, video programming on demand, interactive shopping services, and home telemetry. With these new services, customers might become much more dependent on their local telecommunication system, and any disruption in service would be very undesireable, perhaps, comparable to a power blackout. Thus, reliability issues should assume a much greater importance in planning for future networks. Because of economic considerations in minimizing the number of links, the most common current local network design is a tree configuration. The disadvantage of this design is that any single link failure will disconnect the network. New research is needed to design local access networks that can offer more reliability and resistance to failure (see, for example, Monma and Shallcross (1986) and Groetschel and Monma (1988)). Topologies such as ring networks (which provide two paths between every pair of nodes) might become more common in future local telecommunication systems.

References

Anderson, P. (1988) "Fiber optics for loop applications: A techno-economical analysis", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 213-216.

Ash, G., R. Cardwell, and R. Murray (1981) "Design and optimization of networks with dynamix routing", *Bell System Technical Journal*, **60**, pp. 1787-1820.

AT&T Bell Laboratories (1986) *Engineering and Operations in the Bell System*, Murray Hill, New Jersey.

Bahl, L. R., and D. T. Tang (1972) "Optimization of concentrator locations in teleprocessing networks", *Proceedings of the Symposium on Computer Communication Networks and Teletraffic.*

Balakrishnan, A., T. L. Magnanti, and R. T. Wong (1989a) "A dual ascent procedure for large-scale uncapacitated network design", to appear in *Operations Research.*

Balakrishnan, A., T. L. Magnanti, and R. T. Wong (1989b) "A decomposition algorithm for local access telecommunications planning", in preparation.

Barany, I., J. Edmonds, and L. A. Wolsey (1986) "Packing and covering a tree by subtrees", *Combinatorica*, **6**, pp. 221-233.

Barany, I., T. J. Van Roy, and L. A. Wolsey (1984a) "Uncapacitated lot-sizing: The convex hull of solutions", *Mathematical Programming Study*, **22**, pp. 32-43.

Barany, I., T. J. Van Roy, and L. A. Wolsey (1984b) "Strong formulations for multi-item capacitated lot sizing", *Management Science*, 30, pp. 1255-1261.

Billheimer, J., and P. Gray (1973) "Network design with fixed and variable cost elements", *Transportation Science*, **7**, pp. 49-74.

Boffey, T. B., and A. I. Hinxman (1979) "Solving for optimal network problem", *European Journal of Operational Research*, **3**, pp. 386-393.

Boorstyn, R. B., and H. Frank (1977) "Large-scale network topological optimization", *IEEE Transactions on Communication*, **COM-25**, pp. 29-47.

Boyce, D. E., A. Farhi, and R. Weischedel (1973) "Optimal network problem: A branch-and-bound algorithm", *Enviromental Planning*, **5**, pp. 519-533.

Bulcha, B., L. E. Kodrich, D. B. Luber, W. J. Mitchell, M. A. Schwartz, and F. N. Woomer (1982) "Feeder planning methods for digital loop carrier", *The Bell System Technical Journal*, **61**, No. 9, pp. 2129-2141.

Campbell, L. H. (1988) "The evolution from narrowband to broadband customer access", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 144-148.

Carse, G. D. (1986) "New and future technologies in the local telephone network: The Victoria system", *Proceedings of the IEEE 1986 International Communications Conference*, pp. 410-412.

Chandy, K. M., and T. Lo (1973) "The capacitated minimum spanning tree", *Networks*, **3**, pp. 173-182.

Chandy, K. M., and R. A. Russell (1972) "The design of multipoint linkages in a teleprocessing tree network", *IEEE Transactions on Computers*, **C-21**, pp. 1062-1066.

Ciesielka, A. J., and D. C. Douglas (1980) "Electronics in the suburban and light urban loop networks", *The Bell System Technical Journal*, **59**, No. 3, pp. 417-439.

Ciesielka, A. J., and N. G. Long (1980) "New technology for loops - A plan for the '80's", *IEEE Transactions on Communications*, **COM-28**, No. 7, pp. 923-930.

Coathup, L., J.-P. Poirier, D. Poirier, and D. Kahn (1988) "Fiber to the home - Technology and architecture drives", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 281-285.

Combot, J.-P., and N. Epstein (1979) "The introduction of digital switching to the local network", *IEEE Transactions on Communications*, **COM-27**, No. 7, pp. 1056-1064.

Combot, J.-P., and L. G. Mason (1979) "Optimal switch-concentrator location for a telephone network", *IEEE Transactions on Communications*, **COM-27**, No. 1, pp. 36-43.

Combot, J.-P., M. S. C. Tsui, and R. Weihmayer (1981) "Optimal digital switching introduction into the local network", *IEEE Transactions on Communications*, **COM-29**, No. 10, pp. 1446-1454.

Cornuejols, G., M. L. Fisher, and G. L. Nemhauser (1977) "Location of bank accounts to optimize float: An analytic study of exact and approximate algorithms", *Management Science*, 23, pp. 789-810.

Cornuejols, G., G. L. Nemhauser, and L. A. Wolsey (1989) "The uncapacitated facility location problem" in R. L. Francis and P. Mirchandani (eds.), <u>Discrete Location Theory</u>, Wiley Interscience, to be published.

Cornuejols, G., R. Sridharan, and J. M. Thizy (1987) "A comparison of heuristics and relaxations for the capacitated plant location problem", Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, Pittsburgh.

Dawson, C. S., R. B. Murphy, and E. Wolman (1984) "A history of operations research at Bell Laboratories (1925-1975)", AT&T Bell Laboratories Report.

Dettmer, R. (1985) "The integrated services digital network: Bringing home the bits", *Electronics and Power*, January 1985, pp. 25-28.

Dionne, R., and M. Florian (1979) "Exact and approximate algorithms for optimal network design", *Networks*, **9**, pp. 37-59.

Direlten, H., and R. W. Donaldson (1976) "Topological design of teleprocessing networks using linear regression clustering", *IEEE Transactions on Communications*, **COM-24**, pp. 1152-1159.

Efroymson, M. A., and T. L. Ray (1972) "A branch-and-bound algorithm for plant location", *Operations Research*, **20**, pp. 94-108.

Elken, T. R. (1980) "The application of mathematical programming to loop feeder allocation", *The Bell System Technical Journal*, **59**, No. **4**, pp. 479-500.

Ensdorf, W. E., M. L. Keller, and C. R. Kowal (1988) "Economic considerations of fiber in the loop plant", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 291-296.

Erlenkotter, D. (1978) "A dual-based procedure for uncapacitated facility location", *Operations Research*, **26**, pp. 992-1009.

Esau, L. R., and K. C. Williams (1966) "On teleprocessing system design: Part II", *IBM Systems Journal*, **5**, pp. 142-147.

Fejes-Toth, L. (1959) "Sur la representation d'une population infinie par un nombre fini d'elements", *Acta. Math. Acad. Sci. Hungar.*, **10**, pp. 299-304.

Feldman, E., F. A. Lehrer, and T. L. Ray (1966) "Warehouse location under continuous economies of scale", *Management Science*, **12**, pp. 670-684.

Fisher, M. L. (1981) "The Lagrangian relaxation method for solving integer programming problems", *Management Science*, **27**, pp. 1-18.

Fisher, M. L., and D. S. Hochbaum (1980) "Database location in computer networks", *Journal of ACM*, **27**, pp. 718-735.

Forbes (1986), Novemer 3 issue, pg. 211.

Fortune (1988) "ISDN: The new telephone network", November 1988.

Fortune (1989) "The coming battle", February 13, 1989, pp. 104-107.

Freidenfelds, J., and C. D. McLaughlin (1979) "A heuristic branch-and-bound algorithm for telephone feeder capacity expansion", *Operations Research*, **27**, pp. 567-582.

Garbanati, L. F., and J. R. Palladino (1988) "Fiber optics to the subscriber: Objectives, costs, and technology assessment", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 286-290.

Gavish, B. (1983) "Formulations and algorithms for the capacitated minimal directed tree", *Journal of the ACM*, **30**, pp. 118-132.

Gavish, B., and K. Altinkemer (1986) "Parallel savings heuristic for the topological design of local access tree networks", *Proceedings of IEEE INFOCOM '86*, pp. 130-139.

Geoffrion, A. M., and G. Graves (1974) "Multicommodity distribution system design by Benders decomposition", *Management Science*, **5**, pp. 822-844.

Gibson, A. E., and D. B. Luber (1980) "Critical section methods for loop plant allocation", *The Bell System Technical Journal*, **59**, No. 1, pp. 99-117.

Griffiths, J. M. (1988) Local Telecommunications 2: Into the Digital Era, P. Peregrinus, London, U.K.

Groetschel, M., and C. L. Monma (1988) "Integer polyhedra arising from certain network design problems with connectivity constraints", Report No. 104, Institut for Mathematik, Universitat Augsburg, Augsburg, W. Germany.

Haimovich, M., and T. L. Magnanti (1988) "Extremum properties of hexagonal partitioning and the uniform distribution in Euclidean location", *SIAM Journal of Discrete Mathematics*, **1**, No. 1, pp. 50-64.

Hall, L. A. (1989) *Polyhedral structure of capacitated trees and approximation methods for deterministic scheduling*, Ph. D. Thesis, Operations Research Center, M.I.T., August.

Hall, L. A., and T. L. Magnanti (1988) "Facets for the capacitated minimum spanning tree polytope", Paper presented at the TIMS/ORSA National Meeting, Washington D.C., April.

Helme, M. P., C. Jack, and A. Shulman (1988) "Planning for new services in the local loop", *Proceedings of the International Telecommunications Conference*, **12**, pp. 5.2B.1.1-5.2B.1.12.

Helme, M. P., and T. L. Magnanti (1989) "Designing satellite communication networks using zero-one quadratic programming", *Networks*, **19**, pp. 427-450.

Hoang, H. H., and H. T. Lau (1984) "Optimizing the evolution of digital switching in a local telephone network", *Proceedings of the IEEE International Conference on Communications*, pp. 179-184.

Johnson, D. S., J. K. Lenstra, and A. H. G. Rinnooy Kan (1978) "The complexity of the network design problem", *Networks*, **8**, pp. 279-285.

Kariv, O., and S. L. Hakimi (1979) "An algorithmic approach to network location problems II: The p-medians", *SIAM Journal of Applied Mathematics*, **27**, pp. 539-560.

Kershenbaum, A., and Boorstyn (1975) "Centralized teleprocessing network design", *Proceedings of the National Telecommunication Conference*, pp. 27.11-27.14.

Kershenbaum, A., and W. Chou (1974) "A unified algorithm for designing multidrop teleprocessing networks", *IEEE Transactions on Communications*, **COM-22**, pp. 1762-1772.

Kershenbaum, A., and S. Peng (1986) "Neighbor finding algorithms for CMST calculation", *Proceedings of IEEE INFOCOM '86*, pp. 140-147.

Korkel, M. (1987) "On the exact solution of large-scale simple plant location problems", Working Paper, Deutsche Bundespost, Darmstadt, Federal Republic of Germany.

Konangi, V. K., T. Aidja, and C. R. Dhas (1984) "On the multilevel concentrator location problem for local access networks", *Proceedings of IEEE Globecom '84*, pp. 912-915.

Koontz, W. L. G. (1980) "Economic evaluation of loop feeder relief alternatives", *The Bell System Technical Journal*, **59**, No. 3, pp. 277-293.

Kopp, P. (1986) "A family of solution procedures for the routing problem in medium-term planning of mixed analog-digital transmission networks", *Proceedings of the Third International Networks Symposium* (Networks '86), pp. 138-142.

Kostas, D. J. (1984) "Transition to ISDN - An overview", *IEEE Communications Magazine*, **22**, No. 1, January 1984, pp. 11-17.

Kubat, P. (1985) "Models for allocation of remote switching units to a local area access network", Internal report, GTE Laboratories, Incorporated, Waltham, MA.

Kuehn, A. A., and M. J. Hamburger (1963) "A heuristic program for locating warehouses", *Management Science*, **9**, pp. 643-666.

Lavin, J. S. (1987) "Optimally rehoming local serving offices to a new point of presence", *AT&T Technical Journal*, **66**, pp. 50-54.

Lemke, P., and R. T. Wong (1989) "Characterizations and descriptions of facets for the k-median location problem", in preparation.

Leung, J., and T. L. Magnanti (1989) "Valid inequalities and facets for capacitated plant location problems", *Mathematical Programming*, to appear.

Luna, H. P. L., N. Ziviani, and R. H. B. Cabral (1987) "The telephonic switching centre network problem: Formalization and computational experience", *Discrete Applied Mathematics*, **18**, pp. 199-210.

Magnanti, T. L., and R. T. Wong (1984) "Network design and transportation planning: Models and algorithms", *Transportation Science*, **18**, pp. 1-55.

Mason, L. G. (1983) "Network modernization with capital budget constraints", *Network Planning Symposium* (Networks '83), Brighton, U. K., pp. 5-9.

Mason, L. G. (1984) "An aggregate model for network modernization with constraints", *IEEE Transactions on Communications*, **COM-32**, No. 10, pp. 1073-1079.

Mason, L. G., and J.-P. Combot (1980) "Optimal modernization policies for telecommunications facilities", *IEEE Transactions on Communications*, **COM-28**, No. 3, pp. 317-324.

Mazzei, U., C. Mazzetti, and G. Roso (1988) "Economics of digital carriers and fiber optic systems in subscriber loop", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 133-137.

McGregor, P., and D. Shen (1977) "Network design: An algorithm for the access facility location problem", *IEEE Transactions on Communications*, **COM-25**, pp. 61-73.

Minoux, M. (1987) "Network synthesis and dynamic network optimization", *Annals of Discrete Mathematics*, **31**, pp. 283-324.

Minoux, M. (1989) "Network synthesis and optimum network design problems: Models, solution methods and applications", *Networks*, to appear.

Mirzaian, A (1985) "Lagrangian relaxation for the star-star concentrator location problem", *Networks*, **15**, pp. 1-20.

Monma, C. L., and D. F. Shallcross (1986) "Methods for designing surviveable communication networks", Working paper, Bell Communications Research, Morristown, New Jersey.

Nemhauser, G. L., and L. A. Wolsey (1988) *Integer and Combinatorial Optimization*, John Wiley & Sons, New York.

Oda, T., Y. Wakahara, and M. Furuya (1986) "An efficient decision support system for facility planning of international transmission networks", *Proceedings of the Third International Networks Symposium* (Networks 86), pp. 37-42.

Papadimitriou, C. H. (1981) "Worst case and probabilistic analysis of a geometric location problem", *SIAM Journal on Computing*, **10**, pp. 542-557.

Pirkul, H. (1986) "Location of concentrators in designing local access networks", *Proceedings of IEEE INFOCOM '86*, pp. 148-154.

Rousset, G., and H. Cameron (1986) "Design and optimization of private data networks", *Proceedings of the Third International Network Planning Symposium*, pp. 50-54.

Sa, G. (1969) "Branch and bouhnd and approximate solutions to the capacitated plant location problem", *Operations Research*, **17**, pp. 1005-1016.

Schneider, G. M., and M. N. Zastrow (1982) "An algorithm for the design of multilevel concentrator networks", *Computer Networks*, **6**, pp. 1-11.

Sen, S., R. D. Doverspike, and M. S. Dunatunga (1989) "Unified Facilities Optimizer", Working paper, Systems and Industrial Engineering Department, University of Arizona, Tucson.

Sharma, R. L. (1983) "Design of an economical multidrop network topolocy with capacity constraints", *IEEE Transactions on Communications*, **COM-31**, pp. 590-591.

Shulman, A., and R. Vachani (1988) "A decomposition approach to solving the dynamic outside plant problem", Technical note, GTE Laboratories Incorporated, Waltham, Massachusetts.

Sirbu, M. A., and D. P. Reed (1988) "An optimal investment strategy model for fiber to the home", *Proceedings of the International Symposium on Subscriber Loops and Services* (ISSLS '88), pp. 149-155.

Snelling, R. K., and K. W. Kaplan (1984) "Current and future fiber optics applications - Operating telephone company perspective", *Proceedings of IEEE Globecom '84*, pp. 604-607.

Spielberg, K. (1969) "Algorithms for the simple plant location problem with some side conditions", *Operations Research*, **17**, pp. 85-111.

Tang, D. T., L. S. Woo, and L. R. Bahl (1978) "Optimization of teleprocessing networks with concentrators and multiply connected terminals", *IEEE Transactions on Computers*, **C-27**, pp. 594-604.

Telephony (1985), July 8 issue, pp. 28-29.

The Economist (1987) "A survey of telecommunications: New lines for old", October 17, 1987, pp. 3-34.

Toth, A. G., E. Colombini, P. J. MacLaren, and R. K. Yates (1985) "Fiber in the local exchange network: A planning overview", *Proceedings of the IEEE International Communications Conference*, pp. 520-526.

Ward, J. E., R. T. Wong, P. Lemke, and A. Oudjit (1988) "Properties of the tree k-median linear programming relaxations", Working Paper, Krannert Graduate School of Management, Purdue University, West Lafayette.

White, P. E. (1986) "The broadband ISDN - The next generation telecommunications network", *Proceedings of the IEEE 1986 International Communications Conference*, pp. 385-390.

Woo, L. S., and D. T. Tang (1973) "Optimization of teleprocessing networks with concentrators", *Proceedings of the National Telecommunication Conference*, pp. 37C1-37C5.

Yamamoto, Y., H. Yamamoto, and H. Oikawa (1984) "Access network design with optical fibers", *Proceedings of IEEE Globecom '84*, pp. 637-641.

Figure 1

Hierarchical Structure of Telecommunication Networks

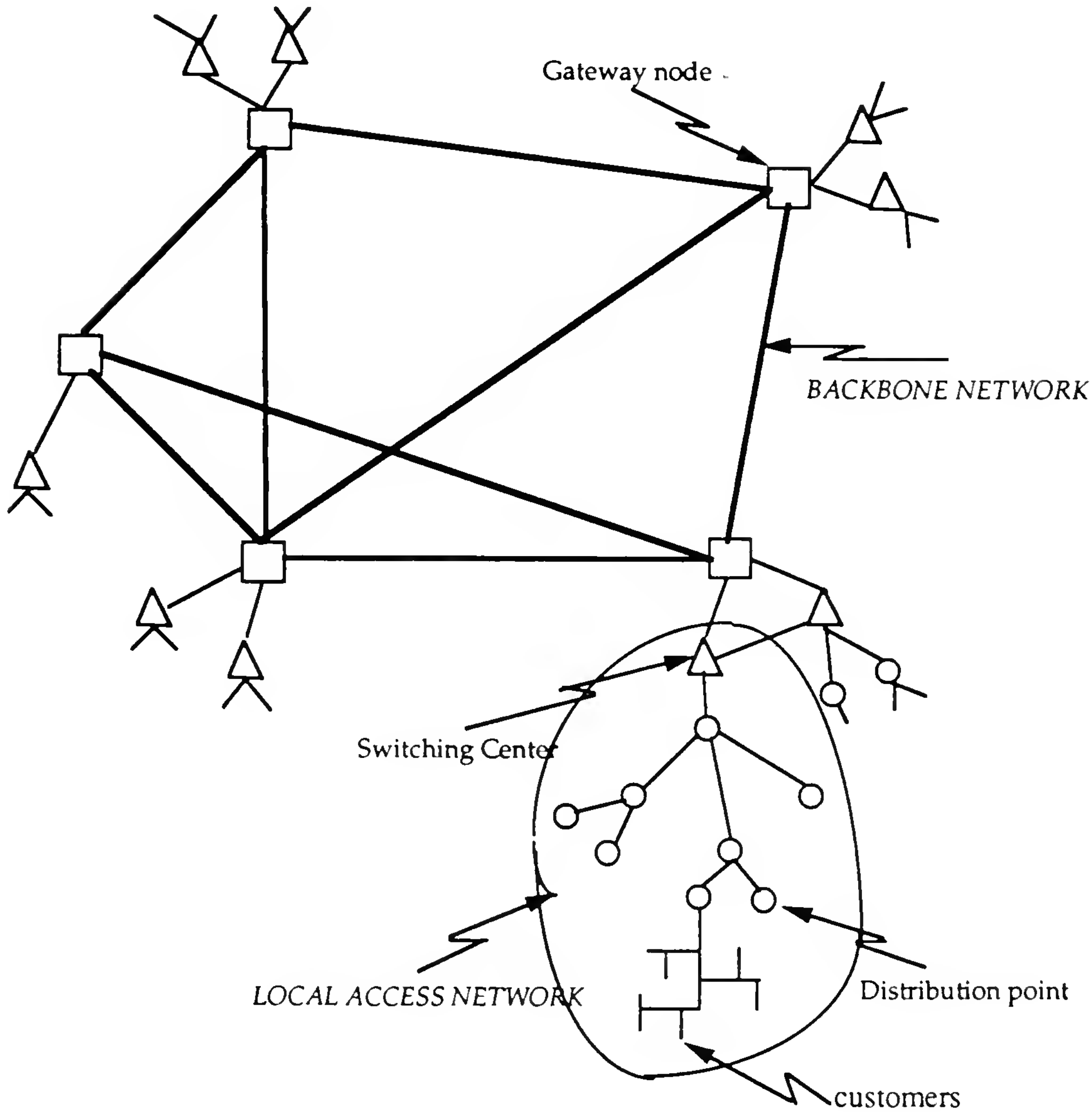


Figure 2

Typical Route of a Local Access Network

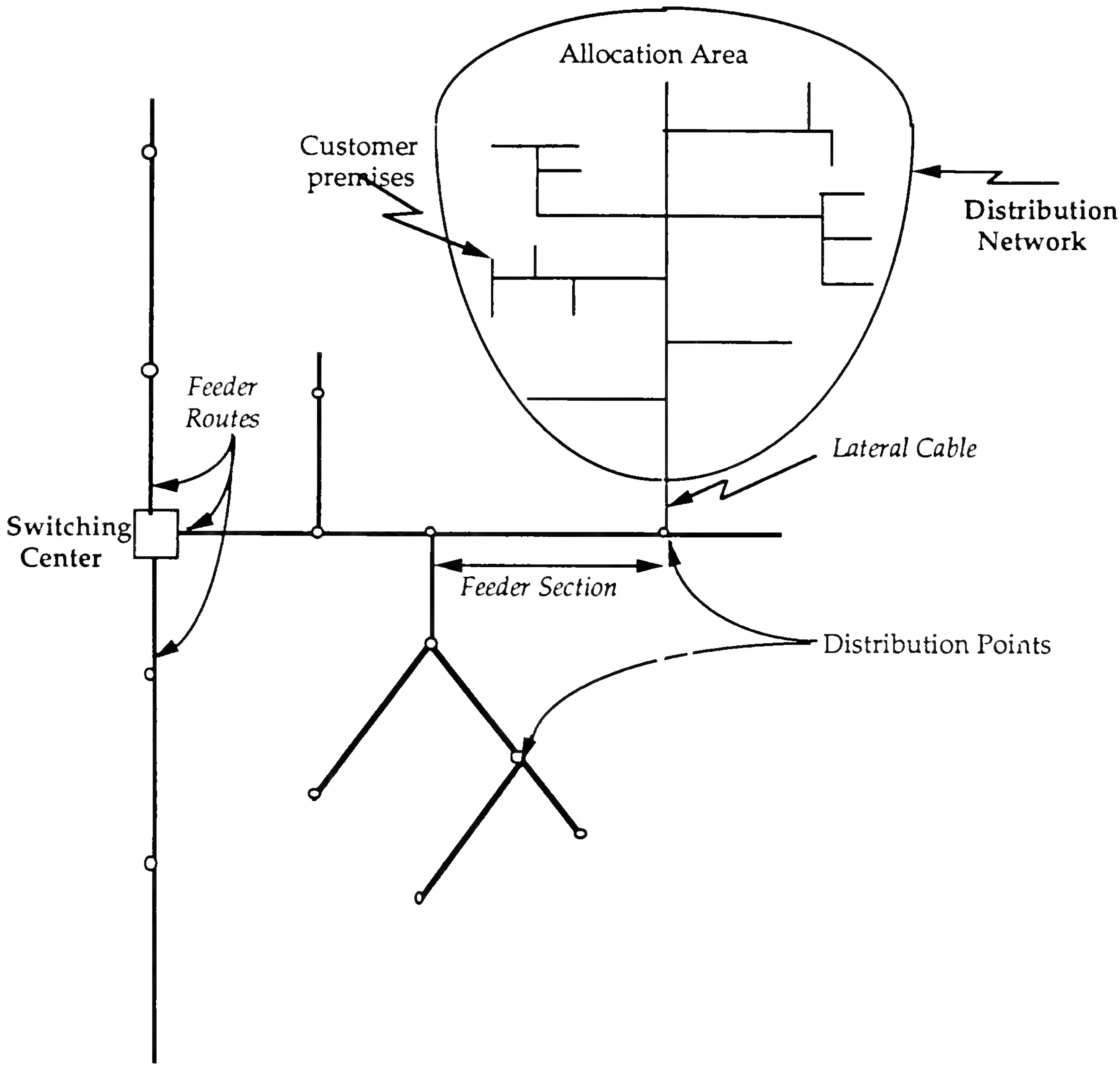
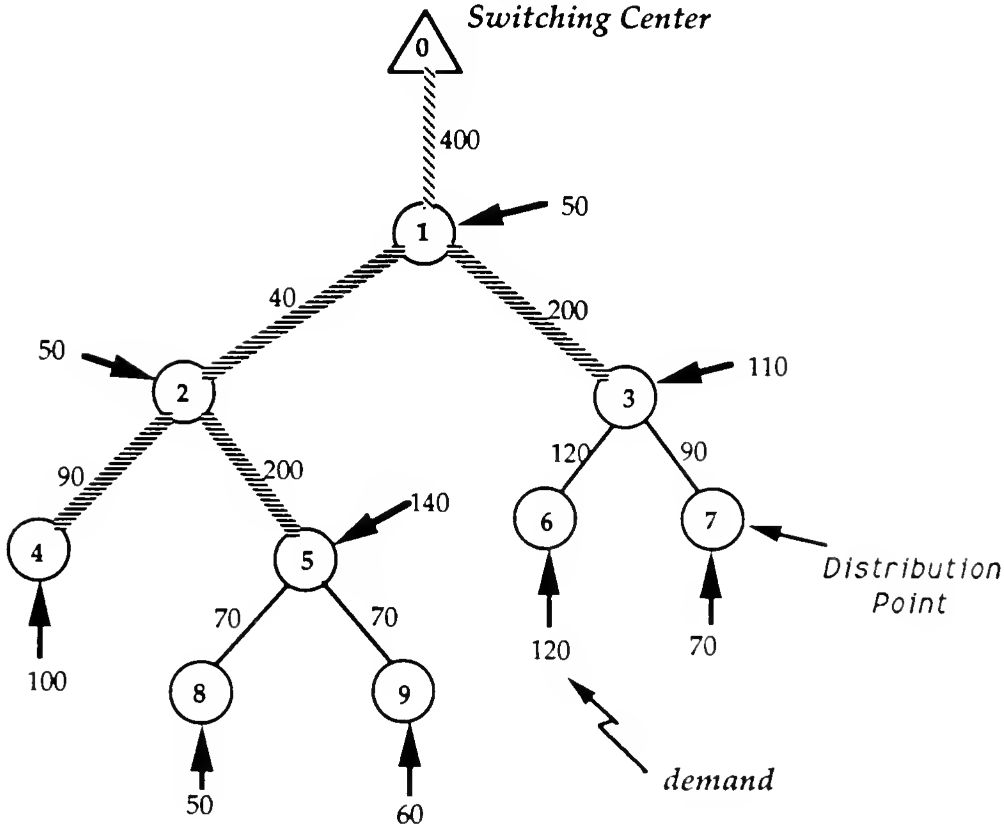


Figure 3A

Local Access Network Planning Example



Numbers on edges denote current cable capacity

BOTTLENECK edges
Cum. demand > Capacity

Figure 3b
Cable Expansion Cost for Local Access Network Planning Example

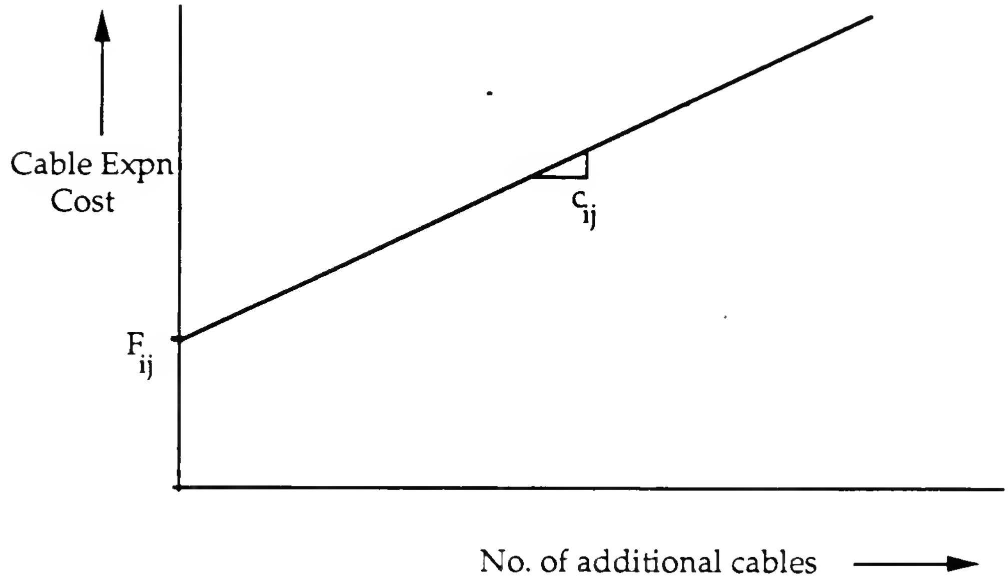


Figure 3c
Processor Cost for Local Access Network Planning Example

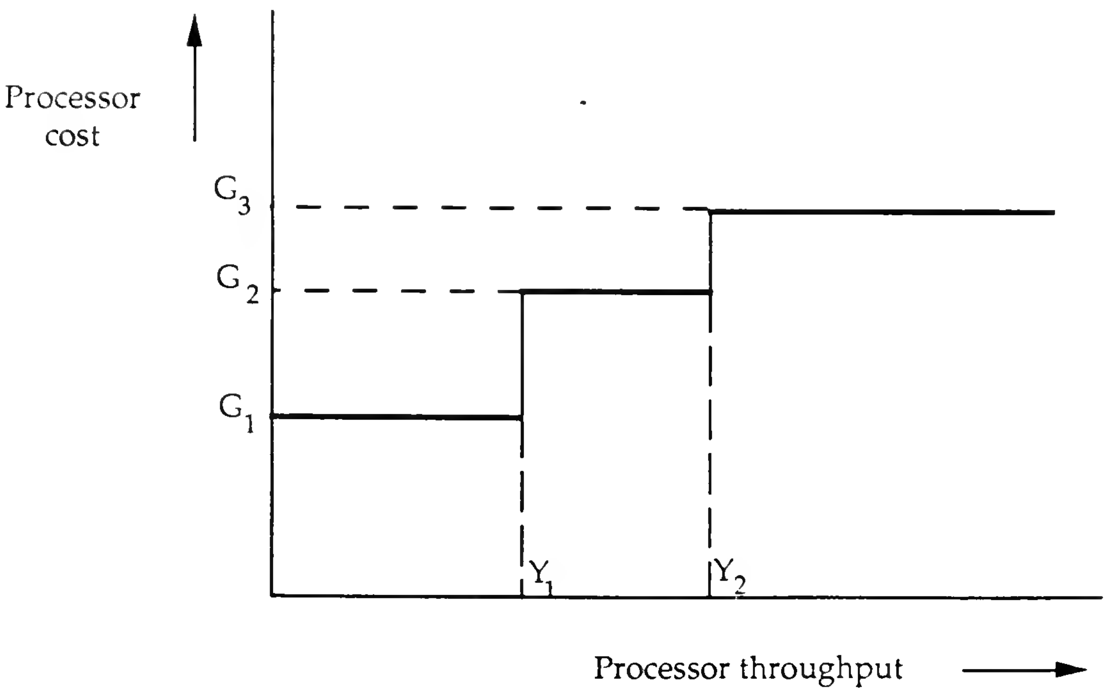
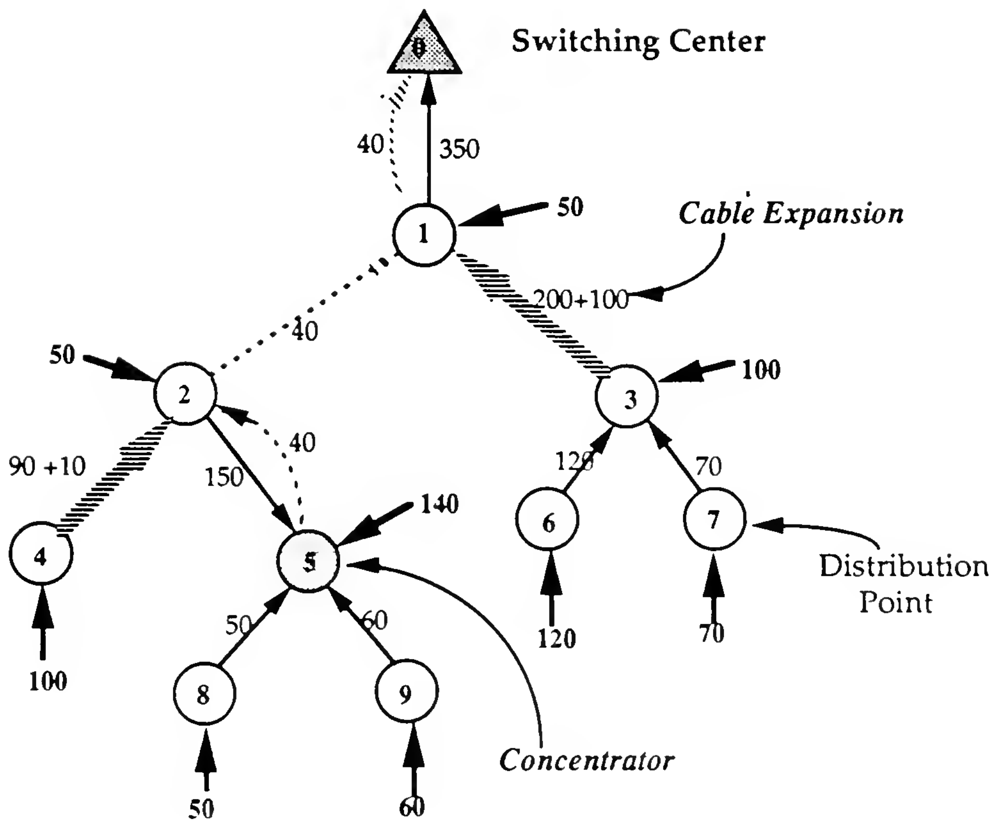


Figure 4

Sample Expansion Plan



·········· Dotted lines show concentrated traffic

Numbers on edges denote number of lines used

Figure 5

Layered Network Representation

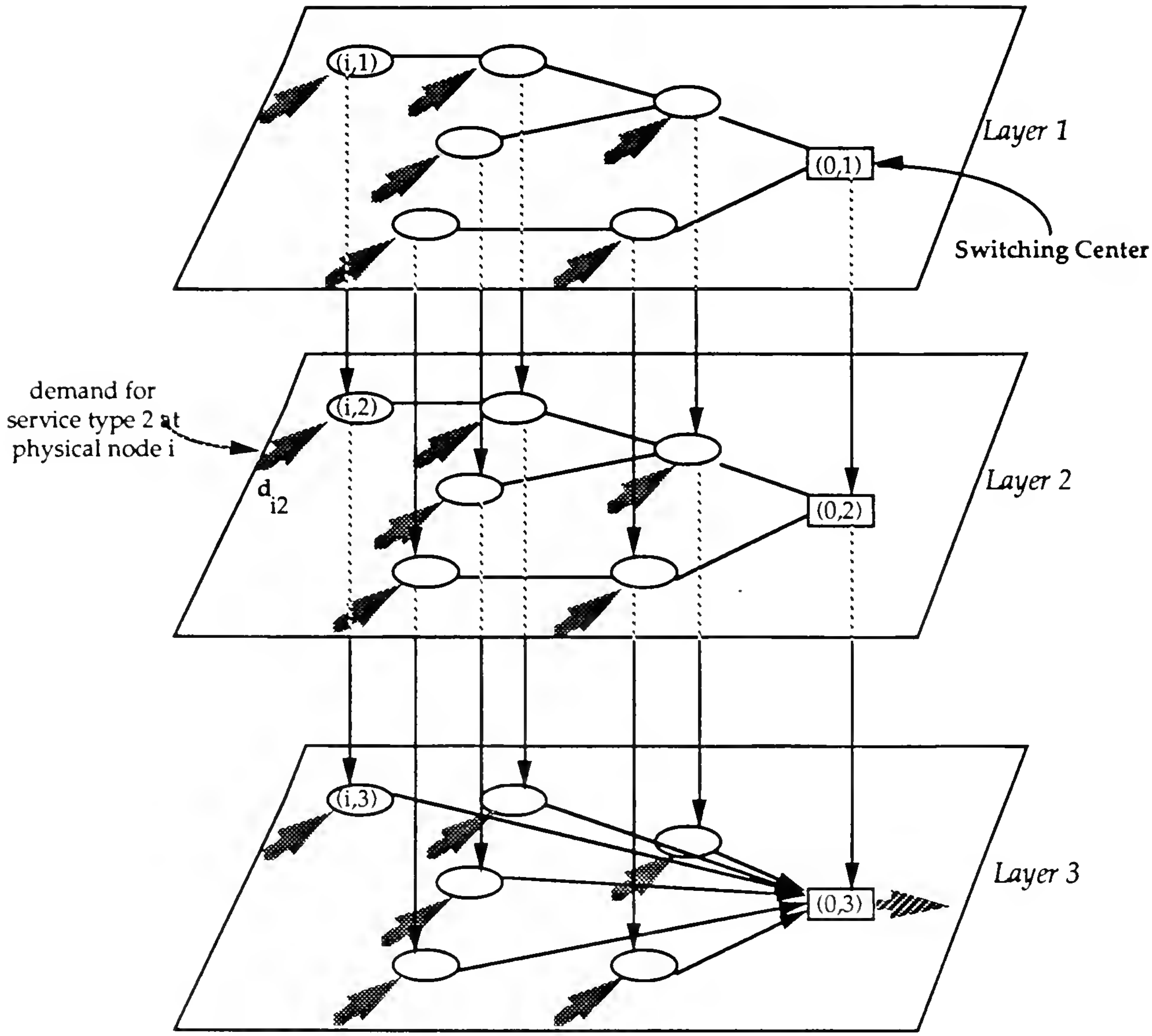


Figure 6
Cable Expansion Cost with Economies of Scale

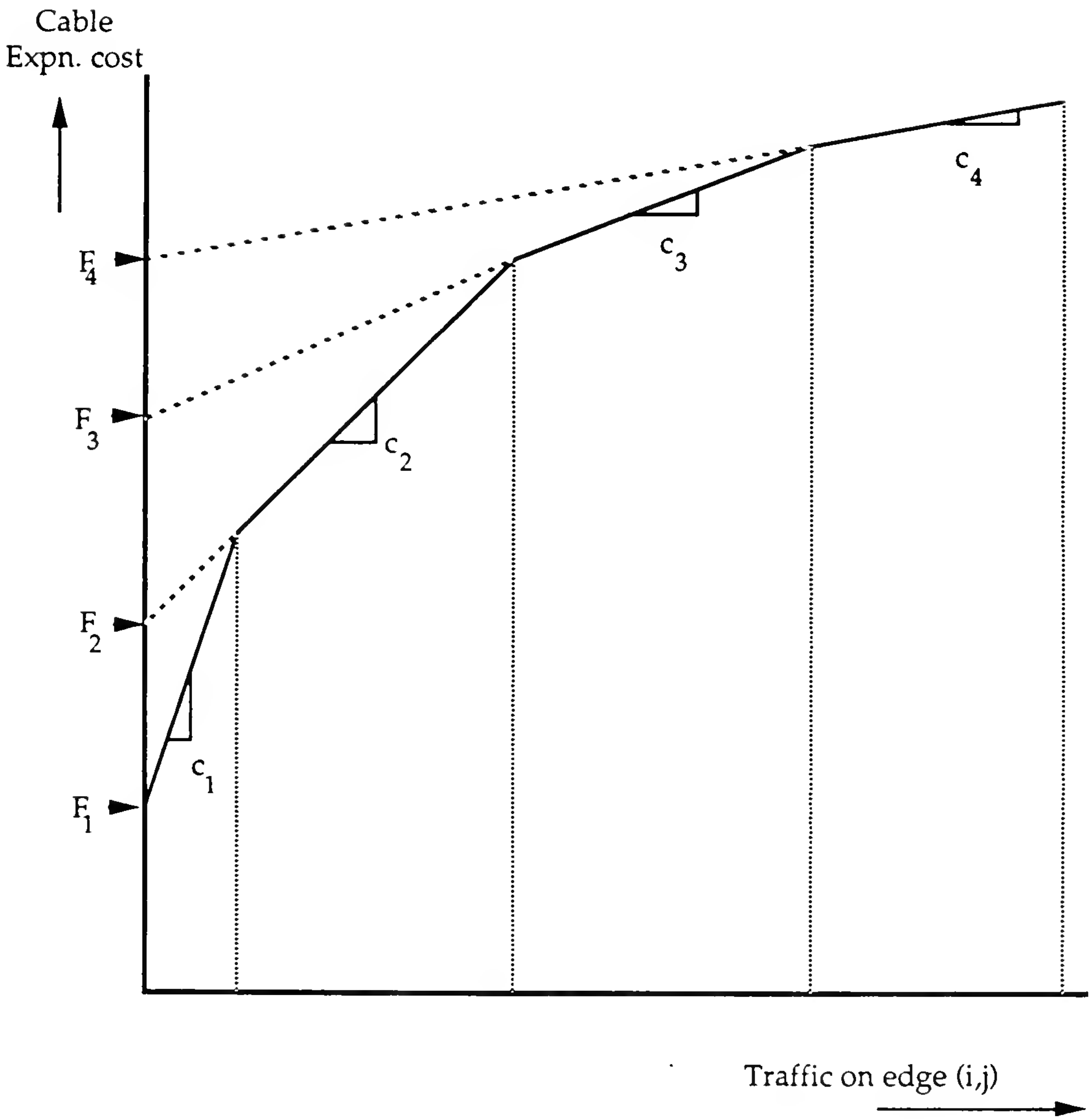


Figure 7A
Convex Cable Expansion Cost

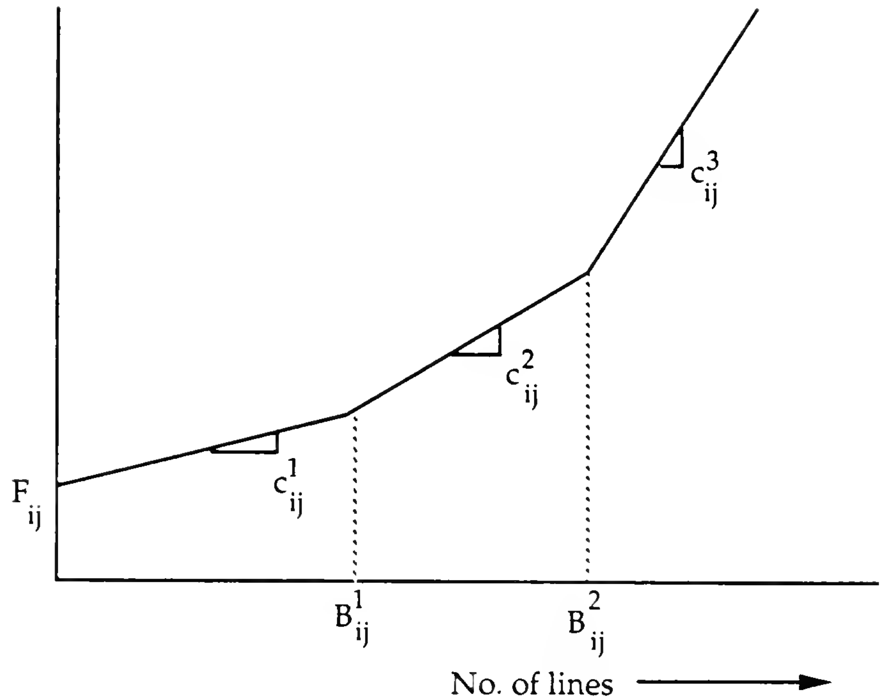


Figure 7B
Equivalent Network Representation

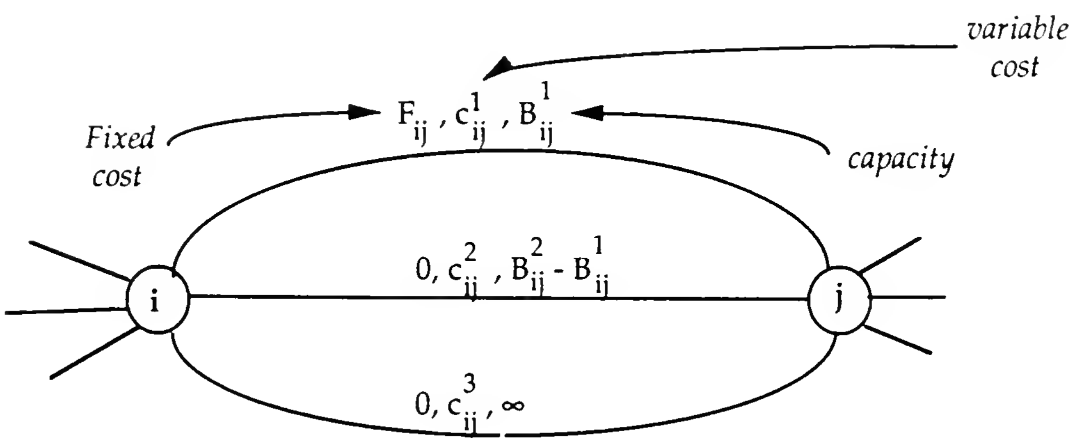


Figure 8A
General Cable Expansion Cost Function

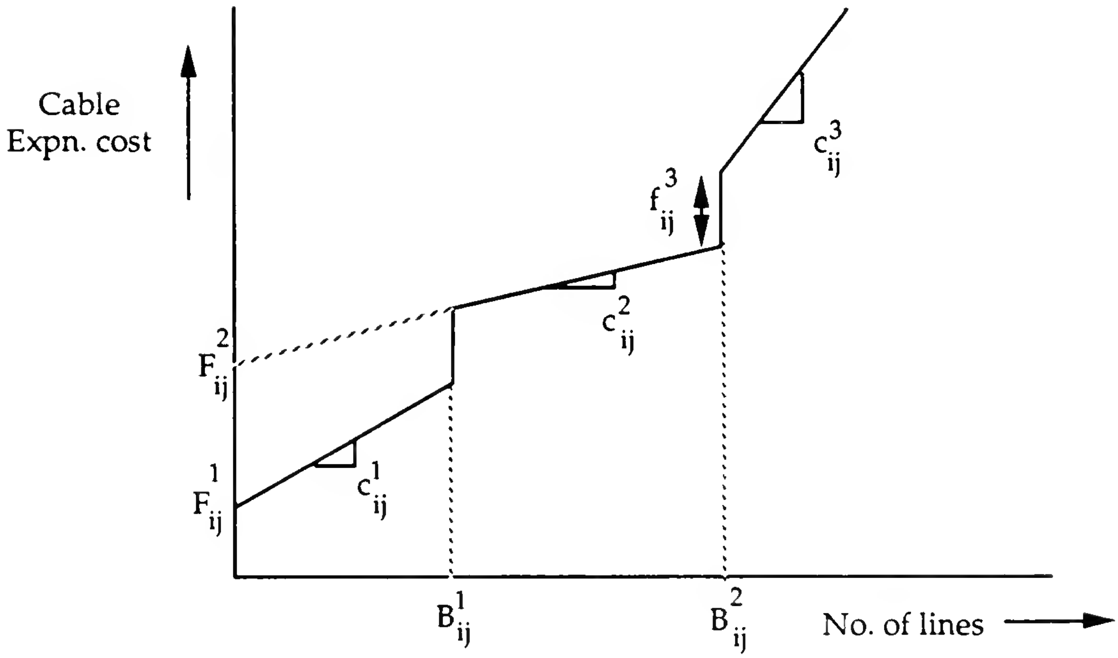


Figure 8B
Equivalent Network Representation

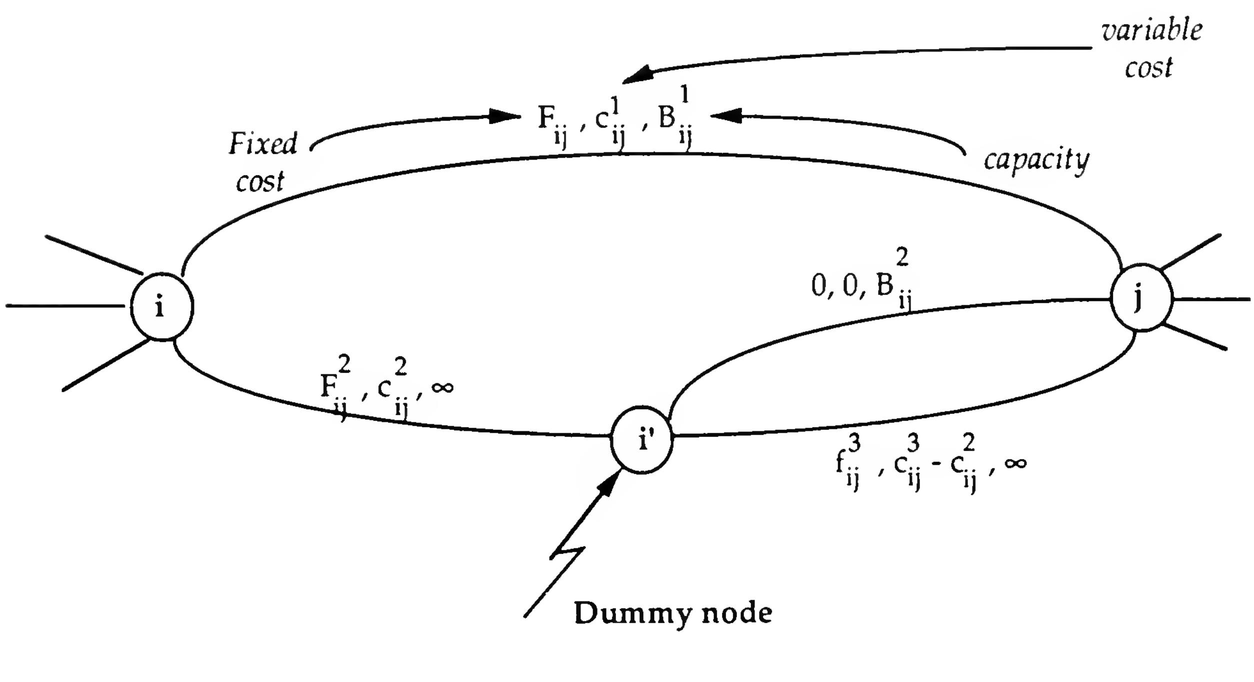


Figure 9
Line Network